Hierarchical Modelling and Forecasting System for Inflation Rate and Volatility

Christoph E. Weiss[a,*], Paul Kattuman[a]

[a] *Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom*

Abstract

Using monthly data that underlies the Retail Prices Index for the UK, we analyse the dynamics of inflation rate and its volatility. We examine patterns in the time-varying covariation among product-level inflation rates that aggregate up to industry-level inflation rates that in turn aggregate up to the overall inflation rate. The aggregate inflation volatility closely tracks the time path of this covariation, which is seen to be driven primarily by the variances of common shocks shared by all products, and by the covariances between idiosyncratic product-level shocks. We formulate a forecasting system that comprises of models for mean inflation rate and its variance, and exploit the index structure of the aggregate inflation rate using the hierarchical time series framework. Using a dynamic model selection approach to forecasting we obtain forecasts that are between 9 and 155 % more accurate than a SARIMA-GARCH(1,1) for the aggregate inflation volatility.

Keywords: Combining forecasts, Decomposition, Disaggregation, Inflation forecasting, Model selection, Multivariate time series, Rule-based forecasting, Uncertainty, Volatility forecasting

1. Introduction

It is widely accepted that inflation and inflation volatility can distort saving, investment and resource allocation decisions (cf. Friedman, 1977; Fischer, 1981). Low and steady inflation rates are the avowed objective of most monetary authorities, many of which have adopted *inflation targeting*, acknowledging the negative consequences of inflation volatility for economic growth and welfare.[1] Notwithstanding their importance, the causal relationship between inflation and inflation volatility is not well understood. The "Friedman-Ball hypothesis" (e.g., Friedman, 1977; Cukierman and Wachtel, 1979; Ball and Cecchetti, 1990) suggests that average inflation rate impacts inflation volatility positively. The "Cukierman-Melzer hypothesis" (e.g., Cukierman and Meltzer, 1986; Holland, 1995) suggests that the causality runs the other way, from inflation volatility to inflation. Kim and Lin (2012) address the reverse causality question using a system of simultaneous equations and panel data for 105 countries, and finds the relationship between inflation and its volatility to be bi-directional, consistent with both hypotheses. This points to the value of a modelling system that can forecast both jointly, which is the basic focus of this paper.

In a comprehensive review of econometric models for inflation, Stock and Watson (2008) compare univariate time series models, backward-looking Phillips curve models, and models with other explanatory variables (e.g., term spread). They conclude that structural (Phillips curve-based) models do not improve upon the

*Correspondence to: Judge Business School, University of Cambridge, Cambridge, CB2 1AG, United Kingdom.
Email address: cw491@cam.ac.uk (Christoph E. Weiss)

[1] Inflation targeting describes a central bank's medium-term goal to reach an explicitly announced target inflation rate through monetary policy actions. In the UK, inflation targeting was first adopted in October 1992 and currently a point inflation target of (annualised) 2 % is applied at all times. For a historical overview of inflation targeting in the UK, see Bean (2003).

forecast accuracy of univariate models overall. "...for the last 15 years, economists have not produced a version of the Phillips curve that makes more accurate inflation forecasts than those from a naïve model" (Atkeson and Ohanian, 2001). Based on estimates from a large set of inflation forecasting models Faust and Wright (2013) agree, concluding that simple models that limit or avoid parameter estimation – e.g., a driftless random walk – are hard to beat.

Univariate time series approaches do not generally distinguish between models for aggregate variables such as the inflation rate, and models for constituent variables such as product-level inflation rates. In modelling aggregates, and particularly the volatility of aggregates, there are potential gains from taking note of *covariance patterns* among the constituents that are nested hierarchically in the aggregate. Indeed, the Great Moderation – the sharply lower aggregate economic volatility from mid-80s to 2007, which is held to be a reason for the poor forecasting performance of structural models (Stock and Watson, 2008) – could be explained in terms of changes in covariance patterns among firms in growth (Comin and Mulani, 2006). Further, in many contexts where the forecast of an aggregate variable (inflation rate) is of interest, aggregation-consistent forecasts of its constituents (product level inflation rates) are also of interest. Finally, given the multiplicity of models, some of which work better than others in time phases that differ in terms of volatility, it would be useful to exploit the potential to by switching between models in order to generate more accurate forecasts.

There is a gap in the literature in terms of modelling systems for inflation and inflation volatility that explicitly consider the way product-level inflation series combine in the aggregate inflation rate and address the above-mentioned issues. In this paper we apply aggregation consistent methods for hierarchical time series (HTS, Athanasopoulos et al., 2009; Hyndman et al., 2011) modelling to obtain forecasts of inflation and inflation volatility. A large number of alternative specifications of HTS models can be used in forecasting, and we offer a dynamic model selection approach that improves the accuracy of forecasts.

2. Inflation rate and its Volatility as Aggregate variables

2.1. Inflation rate

The aggregate inflation rate (Y_t) is constituted as the weighted average of product-level inflation rates, $y_{i,t}$:

$$Y_t = \sum_{i=1}^{N_t} w_{i,t} y_{i,t}$$

where N_t is the number of products in the price index at time t (time-invariant N in this study) and $w_{i,t}$ are the products' weights (share) in the price index at time t.[2] A second level decomposition of product-level inflation rates, into common, industry, and idiosyncratic parts, can be helpful in understanding the inflation data generating process.

$$\begin{aligned} y_{i,t} &= c_t + I_{i,t} + \epsilon_{i,t} \\ Y_t &= \sum_{i=1}^{N} w_{i,t} y_{i,t} = \sum_{i=1}^{N} w_{i,t} \left(c_t + I_{i,t} + \epsilon_{i,t} \right) \end{aligned}$$

[2] In this study, the aggregate inflation rate is the month-to-month growth rate of the Retail Price Index (RPI). The product-level inflation rates, $y_{i,t}$, are the month-to-month growth rates of the respective product price indices, $p_{i,t}$, calculated as: $y_{i,t} = \frac{p_{i,t} - p_{i,t-1}}{(p_{i,t} + p_{i,t-1})/2}$. This preserves seasonality in the series that we take to out modelling exercises, rather than difference it out. This growth rate estimator is symmetric about zero, and bounded, allowing the treatment of entries, exits, and continuers on the same footing (Comin and Mulani, 2004; Davis et al., 2006). It allows for consistent aggregation, and is identical to log differences up to a second-order Taylor Series expansion. See Davis et al. (2006) and references therein for discussions of the appeal of this estimator.

where c_t is the part of a product's inflation rate that is shared with all products, attributable to common shocks – (weighted) average over all product price growth rates; $I_{i,t}$ is the part of i^{th} product's inflation rate that it shares with products in the same industry, but not with products in other industries – (weighted) excess growth rate of the price the product, relative to the industry that product i belongs to; and $\epsilon_{i,t}$ is the excess inflation rate for product i relative to the sum of the common and the industry parts relating to it – the residual.

Fig. 1 illustrates this decomposition for 'Oil and Other Fuels'. It is obvious that the idiosyncratic part is the main driver of the product-level inflation rate.

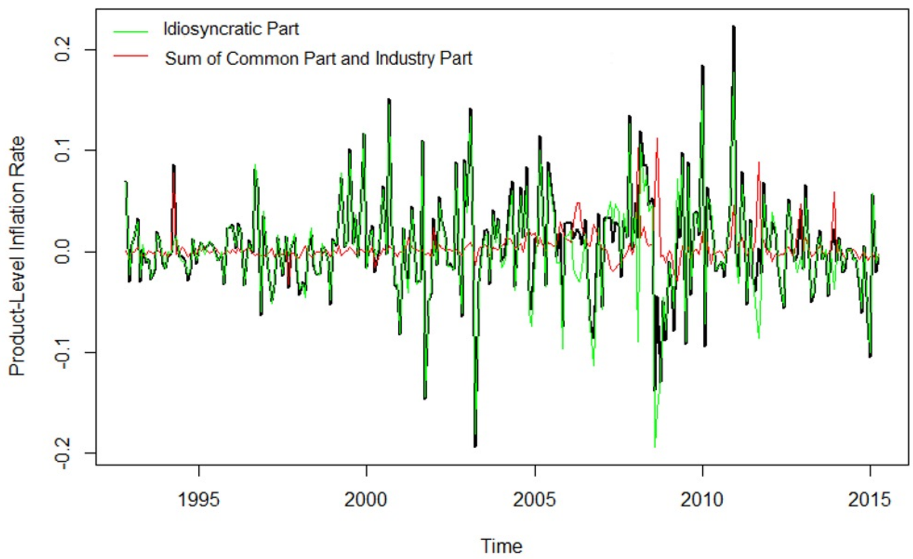

Fig. 1. Product-Level Inflation Rate Decomposition: Oil and Other Fuels.

2.2. *Variance of Inflation rate*

To measure volatility, we use time-varying inflation volatility estimates that are not contingent on any specific model. These are non-parametric estimates from the ex-post sample-paths of inflation over *rolling windows* of time. The simplest such is the simple moving average (SMA) variance. For a generic variable Y, over the chosen discrete time interval $m > 0$:

$$\hat{\sigma}^{2}_{Y_t}{}^{roll} = \frac{\sum_{\tau=t-m}^{t-1}(Y_\tau - \hat{Y}_t)^2}{m}$$

where $\hat{Y}_t$ is the estimate of $E(Y_t)$, the time-varying expectation, the mean of Y over the interval (Comin and Mulani, 2006).[3] Conceptually, instantaneous volatility is the limiting value as $m \to 0$ in the $[t-m, t]$ interval, with m determining the bias-variance tradeoff of the estimator (larger values of m reducing variance, but increasing bias) in the interpretation of $\hat{\sigma}^{2}_{Y_t}{}^{roll}$ as estimate of the 'current' variance of Y_t. The exponentially weighted moving average (EWMA) variance estimator which we use, is more appealing in that it places higher weights on more recent observations:

[3] This volatility estimate for time t is one-sided with respect to t rather than symmetric about t. This makes it suitable for forecasting.

$$\hat{\sigma}^{2\ roll}_{Y_t} = \sum_{\tau=t-m}^{t-1} \alpha_\tau (Y_\tau - \hat{Y}_t)^2$$

where weight scheme is: $\alpha_{\tau+1} = \lambda\alpha_\tau$, with $\lambda \in [0,1]$ the decay factor. A higher λ indicates slower decay, i.e., indicates strong persistence in volatility.[4]

The (rolling window EWMA) variance of the aggregate inflation rate, $\hat{\sigma}^{2\ roll}_{Y_t}$ can be written in terms of weighted (rolling window) estimators of product-level inflation rate variances and covariances. An exact decomposition in terms of a "variance component" (VC) and a "covariance component" (CC) is straightforward (Comin and Mulani, 2006):[5]

$$\hat{\sigma}^{2\ roll}_{Y_t} = \underbrace{\sum_i w_{i,t}^2 \hat{\sigma}^{2\ roll}_{y_{i,t}}}_{\text{VC}} + \underbrace{\sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{\ roll}_{y_{i,t},y_{j,t}}}_{\text{CC}}$$

Our focus in the volatility modelling part is on identifying the dynamic patterns in the way time-varying variances and covariances feed into the *volatility of the aggregate*, in order to forecast aggregate volatility better. Characterizing time series dependencies in the variance component and the covariance component is a useful step in this. We work with estimates of product-level variances $\hat{\sigma}^{2\ roll}_{y_{i,t}}$, and covariances, $\hat{\sigma}^{roll}_{y_{it},y_{jt}}$ that are EWMA-smoothed.[6]

$$\begin{aligned}
\hat{\sigma}^{2\ roll}_{y_{i,t}} &= \sum_{\tau=t-m}^{t-1} \alpha_\tau (y_{i,\tau} - \hat{y}_{i,t})^2 \\
\hat{\sigma}^{\ roll}_{y_{i,t},y_{j,t}} &= \sum_{\tau=t-m}^{t-1} \alpha_\tau (y_{i,\tau} - \hat{y}_{i,t})(y_{j,\tau} - \hat{y}_{j,t})
\end{aligned}$$

The time-varying volatility can be further decomposed into common, industry, and idiosyncratic parts. The Variance Component (VC) is:

$$\widehat{VC}_t = \underbrace{\sum_i w_{i,t}^2 \hat{\sigma}^{2\ roll}_{c_t}}_{\substack{\text{VC} \\ \text{(Var Common)}}} + \underbrace{\sum_i w_{i,t}^2 \hat{\sigma}^{2\ roll}_{I_{i,t}}}_{\substack{\text{VC} \\ \text{(Var industry)}}} + \underbrace{\sum_i w_{i,t}^2 \hat{\sigma}^{2\ roll}_{\epsilon_{i,t}}}_{\substack{\text{VC} \\ \text{(Var Idio.)}}} +$$

$$+ \underbrace{2\sum_i w_{i,t}^2 \hat{\sigma}^{\ roll}_{c_t,I_{i,t}}}_{\substack{\text{VC} \\ \text{(Cov Common/Ind)}}} + \underbrace{2\sum_i w_{i,t}^2 \hat{\sigma}^{\ roll}_{c_t,\epsilon_{i,t}}}_{\substack{\text{VC} \\ \text{(Cov Common/Idio.)}}} + \underbrace{2\sum_i w_{i,t}^2 \hat{\sigma}^{\ roll}_{I_{i,t},\epsilon_{i,t}}}_{\substack{\text{VC} \\ \text{(Cov Ind/Idio.)}}}$$

[4] In asset pricing, $E(Y_t)$, which is estimated as $\hat{Y}_t$, is usually assumed to be zero; this collapses the EWMA model to a zero-intercept IGARCH(1,1):

$$\hat{\sigma}^2_{Y_t} = \lambda\sigma^2_{t-1} + (1-\lambda)Y^2_{t-1}$$

Another appealing feature favouring the use of EWMA is the existence of variants that take note of distributional divergence from normality, which allows its use with great flexibility in our framework. See Lucas and Zhang (2016) for a review of robust EWMA, skewed EWMA, and fat-tailed skewed EWMA and their corresponding GARCH and GAS/DCS models.

[5] See Comin and Mulani (2004), p.13 for a derivation of the variance identity.

[6] It is worth considering how inflation targeting can be incorporated into the model. One possibility is to assume that the expected inflation rate $E(Y_t) = \mu$ is the target inflation rate. The EWMA model under this assumption is:

$$\hat{\sigma}^2_{Y_t} = \lambda\sigma^2_{t-1} + (1-\lambda)Y^2_{t-1} - 2(1-\lambda)Y_{t-1}\mu + (1-\lambda)\mu^2$$

The volatility estimate is now a function of lagged inflation rate and target inflation rate as well.

The Covariance Component (CC) is:

$$\widehat{CC}_t = \underbrace{\sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{2\ roll}_{c_t}}_{\substack{\text{CC} \\ \text{(Var Common)}}} + \underbrace{2 \sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{roll}_{c_t, I_{j,t}}}_{\substack{\text{CC} \\ \text{(Cov Common/Ind)}}} +$$

$$+ \underbrace{2 \sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{roll}_{c_t, \epsilon_{j,t}}}_{\substack{\text{CC} \\ \text{(Cov Common/Idio.)}}} + \underbrace{\sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{roll}_{I_{i,t}, I_{j,t}}}_{\substack{\text{CC} \\ \text{(Cov Ind/Ind)}}} +$$

$$+ \underbrace{2 \sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{roll}_{I_{i,t}, \epsilon_{j,t}}}_{\substack{\text{CC} \\ \text{(Cov Ind/Idio.)}}} + \underbrace{\sum_i \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\sigma}^{roll}_{\epsilon_{i,t}, \epsilon_{j,t}}}_{\substack{\text{CC} \\ \text{(Cov Idio./Idio.)}}}$$

We proceed to hierarchical time series modelling of the inflation system comprising mean inflation rate and inflation volatility, based on the above decompositions. We examine dynamic patterns in and between the variance and the covariance components, to distinguish high volatility / low volatility periods. We then proceed to forecast aggregate inflation rate and volatility, and compare results with extant volatility forecasting approaches.

3. Data and Descriptive analysis

3.1. Data: UK Retail Prices Index (RPI)

RPI is a long-standing measure of inflation in the UK, though it is no longer designated an official National Statistic (Office for National Statistics, 2013). RPI measures inflation with reference to the cost of a "representative basket of goods and services bought by consumers within the UK". It is calculated from the same basic price data as the CPI, and uses similar methodology in compiling and aggregating the constituent price indices. RPI covers 85 products.[7] Monthly data from January 1987 has been published by the Office for National Statistics. The weights are updated at the beginning of each year using the information on household spending. They are relatively unchanging, and may be considered exogenous.[8]

3.2. Descriptive Analysis

Inflation rate. Fig. 2 presents stylised facts of the monthly growth rate of the RPI and its volatility. The most notable features are the increased stability at a low inflation level after the introduction of inflation targeting (October 1992) and increased inflation volatility during recessions. The behaviour of inflation does appear to be driven by changes in government or changes in the governance of the Bank of England. In the analysis that follows, we focus on the period after introduction of inflation targeting.

Inflation Volatility. We begin with the decomposition of rolling sample variance of the aggregate inflation rate into its variance and covariance components (Fig. 3). Two evident phases of high volatility appear to differ sharply from each other. In the early 90s, the recession saw high inflation and high volatility.

[7]For a full documentation of the representative products in the RPI, see Annex C in Beeson (2016).

[8]The Herfindahl-Hirschman index (HHI) for the RPI lies in the range of 0.0210 to 0.0289 for the entire sample period from 1987 to 2015. For 85 items, a HHI of 0.0117 would indicate equal weighting, while a value of 1 would indicate concentration in a single product, suggesting that the RPI is relatively unconcentrated over the entire observation period.

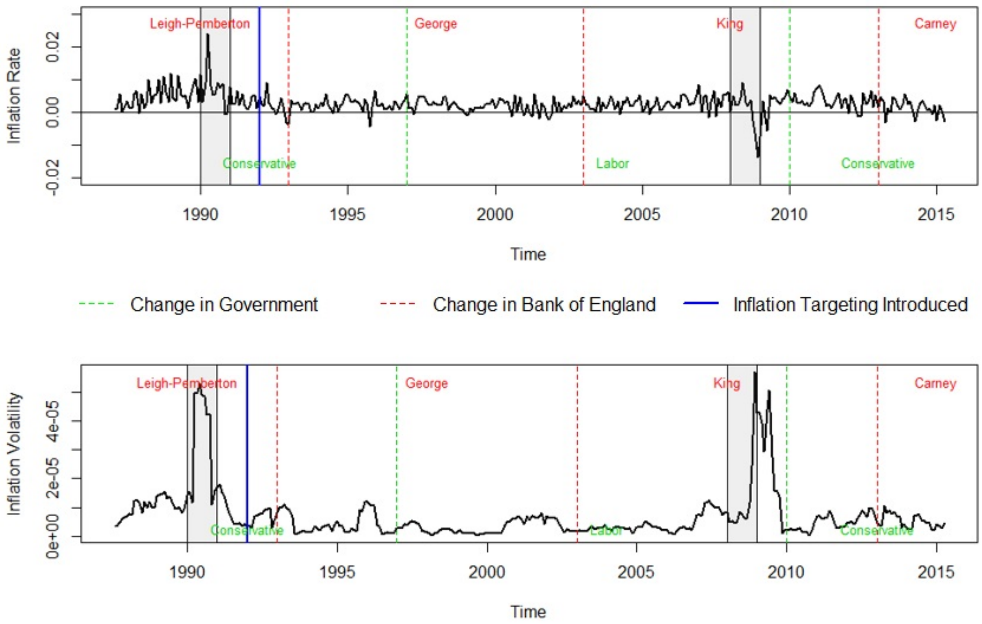

Fig. 2. Stylised Facts: UK Inflation – Impacts of Recessions and Monetary Policy.

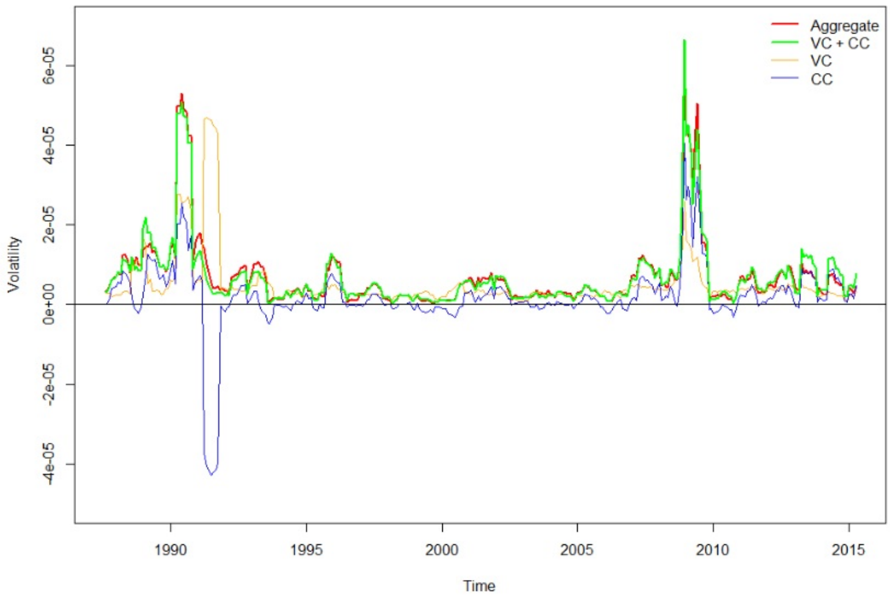

Fig. 3. Decomposition of Aggregate Inflation Variance into Variance (VC) and Covariance (CC) Components.

The Great Recession saw negative inflation rate and high volatility. Both the variance and the covariance components differed in behaviour immediately following the first and the second high volatility periods. A more detailed product-level analysis sheds some light on the underlying dynamics at play during these recession phases:

- The spikes in the Variance Component (the contribution of product-level variances to aggregate variance) are mostly due to a single product. The increased variance from April to October 1990 is caused by 'Council Tax and Rates' (responsible for 75 - 83 % of the Variance Component in this period); the increased value from April to October 1991 is caused by the same item (69 - 74 % of VC in this period), the spike in VC in December 2008 is due to the item 'Mortgage Interest Payments' that contributed 63 % that month.[9]

- Examining the Covariance Component, it is evident that product-level inflation rates co-move more strongly during recessions. The spikes in the CC are also linked to the items that cause increases in the VC: The main contributors to the positive spike in the CC from April to October 1990 are the positive covariances between 'Council Tax and Rates' and 'Rent', and between 'Council Tax and Rates' and 'Petrol and Oil' – all of which had positive product-level inflation rates. The main contributors to the CC from April to October 1991 were the negative covariances between 'Council Tax and Rates' (which showed a negative product-level inflation rate) with 'Beer on Sales', 'Rent' and 'Petrol and Oil' (which showed positive product-level inflation rates); as well as the positive covariance between 'Council Tax and Rates' and 'Mortgage Interest Payments' (which also showed a negative product-level inflation rate). The main contributors to the CC from December 2008 to July 2009 (note that the CC remains high for longer than the VC) are the positive covariances of 'Mortgage Interest Payments' with 'Gas' and with 'Petrol and Oil' (all of which showed negative product-level inflation rates).

In recessions, both the variance and covariance components of aggregate volatility tend to increase – with elevated variances of product-level inflation rates and elevated covariances across products. In the immediate aftermath of the high-volatility phase the covariance component is negative, while the variance component remains high. While the variance and covariance components are similar in magnitude, the dynamics of the covariance component clearly drive the path of aggregate volatility, in particular in the aftermath of recessions. It is clearly important to understand covariation among product-level inflation rates.

Decomposition of the Variance and Covariance components of Volatility. We turn to the decomposition of the variance component, using the breakdown of the product-level inflation rate into common, industry, and idiosyncratic parts (Fig. 4). It appears that product-level inflation rate variances are mainly driven by the variances of their idiosyncratic part. This result seems reasonable given our earlier finding that the idiosyncratic part is the dominant driver of product-level inflation rates and that the spikes in the VC are driven by high variances of individual products, rather than an increased variance of all products in the economy or an industry.

Fig. 5 shows the decomposition of the covariance component (again using the breakdown into common, industry, and idiosyncratic at the product level). The variance of the common part appears to be the main driver of this component. This is intuitive, since the common part is designed to capture positive co-movement of the entirety of product-level inflation rates. Correspondingly, covariances between idiosyncratic parts capture a good part of the negative co-movement between product-level inflation rates. These patterns are consistent with the patterns found in the product-level analysis of recessions – during these phases, the CC is driven by a combination of stronger co-movement between all products, as well as covariances between individual products that are most responsible for the high aggregate volatility.

[9] The time series of the monthly inflation rates corresponding to these items can readily be accessed on the ONS website: www.ons.gov.uk/economy/inflationandpriceindices/timeseries/sgpr/mm23 (Council Tax and Rates), www.ons.gov.uk/economy/inflationandpriceindices/timeseries/sgpn/mm23 (Mortgage Interest Payments). The specific drivers in play are likely to have been the short-lived poll tax in 1990 and 1991 (e.g., Ridge and Smith, 1991), and the sub-prime crisis in 2008 (e.g., Galati et al., 2011).

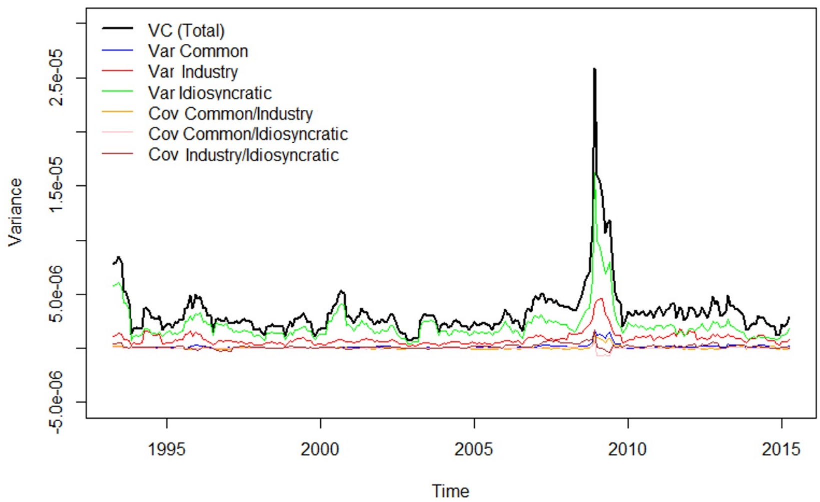

Fig. 4. Decomposition of Variance Component: Common, industry, and Idiosyncratic Part (Product-Level)

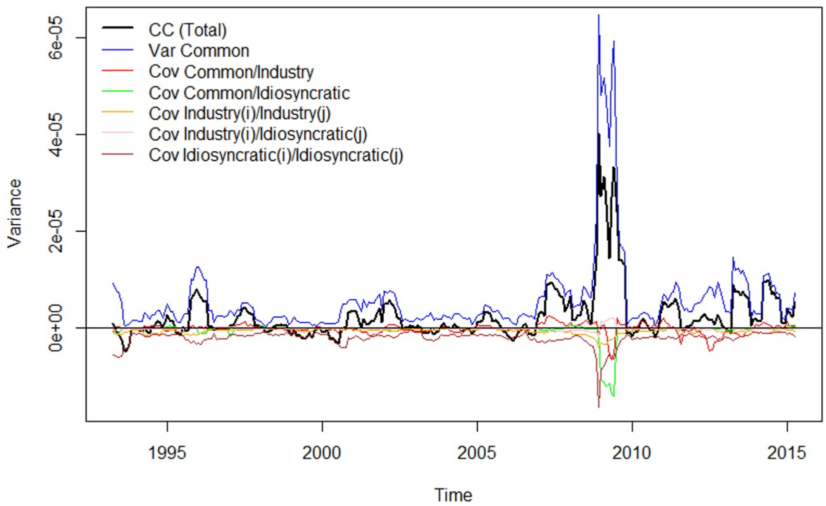

Fig. 5. Decomposition of Covariance Component: Common, industry, and Idiosyncratic Parts (Product-Level).

The above summary points to the potential for understanding phases of high aggregate volatility in terms of the dynamics in product-level inflation rates, paying attention to combining aggregate shocks (variance of common part) and industry/idiosyncratic shocks (covariances between industry and idiosyncratic parts).

Quah (1994) sounded an early warning of the potential fallacy of composition involved in modelling dynamics of macro aggregates ignoring the dynamic behaviour of disaggregates. The question whether aggregate volatility can be caused by shocks at microeconomic levels has been subject of much recent attention (e.g., Comin and Mulani, 2006; Gabaix, 2011; Carvalho and Gabaix, 2013). Abadir and Talmain (2002) have argued that common shocks are a more potent driver of aggregate fluctuations than idiosyncratic shocks. The disaggregated analysis of inflation suggests that aggregate volatility is driven by a combination of common shocks (through the CC) and idiosyncratic shocks (through the VC). We now turn to the question of using disaggregated information to forecast aggregate inflation and its volatility.

4. Forecasting Inflation Rate and Volatility: HTS

The issue of optimally forecasting contemporaneously aggregated variables – in particular the attempt of finding an optimal level of disaggregation for that purpose – is an ongoing debate in forecasting research (Hendry and Hubrich, 2006; Chen and Boylan, 2007, 2009). The approaches include: forecasting the aggregate using only aggregate information, forecasting the aggregate by aggregating forecasts of disaggregates, and forecasting the aggregate using information on disaggregates.

It is well established that if the data-generating process is known to the forecaster, then a procedure that aggregates the forecasts of disaggregates always outperforms direct forecasting of the aggregate, as the disaggregate information can be used optimally. But when the data-generating process is not known as is common, the uncertainties in specification and estimation make the relative efficacies of the two approaches an empirical question. The Hierarchical Time Series (HTS) procedure of Athanasopoulos et al. (2009) is suitable for addressing this empirical question.

The starting point is to write the inflation rate as a hierarchical time series, noting that the index weights feature in the aggregation procedure. The decomposition involved is multi-stage, involving 1 aggregate series, 15 industry-level inflation rates, 85 product-level inflation rates, and 255 inflation rates that correspond to the decomposition of product-level inflation rates into common, industry, and idiosyncratic parts). The total number of time series in a hierarchy with K levels is $n = 1 + n_1 + \cdots + n_K$, where n_i is the number of series at level i of the hierarchy (356 individual time series in our exercise).

Expressed in matrix form, the hierarchical time series ($\mathbf{Z}_t$) is obtained by applying a summation matrix ($\mathbf{S}$, of order $n \times n_K$) to aggregate the base-level series ($\mathbf{z}_{K,t}$). For the inflation rate, it is necessary to extend the simple aggregation in Athanasopoulos et al. (2009), to weighted aggregation.[10] To do this the simple summation matrix can be rewritten to contain the corresponding index weights of the bottom-level series.

As an illustration, consider 9 products (labelled 1 to 9) that aggregate into 4 industries (AA, AB, BA, BB), which then aggregate into core inflation (A) and non-core inflation (B), and finally into inflation (Y).[11] The resulting hierarchical time series, can be expressed as follows:

$$\underbrace{\begin{pmatrix} P_{Y,t} \\ P^w_{y_A,t} \\ P^w_{y_B,t} \\ P^w_{y_{AA},t} \\ P^w_{y_{AB},t} \\ P^w_{y_{BA},t} \\ P^w_{y_{BB},t} \\ P^w_{y_1,t} \\ P^w_{y_2,t} \\ P^w_{y_3,t} \\ P^w_{y_4,t} \\ P^w_{y_5,t} \\ P^w_{y_6,t} \\ P^w_{y_7,t} \\ P^w_{y_8,t} \\ P^w_{y_9,t} \end{pmatrix}}_{\mathbf{Z}_t} = \underbrace{\begin{pmatrix} w_{1,t} & w_{2,t} & w_{3,t} & w_{4,t} & w_{5,t} & w_{6,t} & w_{7,t} & w_{8,t} & w_{9,t} \\ w_{1,t} & w_{2,t} & w_{3,t} & w_{4,t} & w_{5,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{6,t} & w_{7,t} & w_{8,t} & w_{9,t} \\ w_{1,t} & w_{2,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_{3,t} & w_{4,t} & w_{5,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{6,t} & w_{7,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{8,t} & w_{9,t} \\ w_{1,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{2,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_{3,t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{4,t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{5,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{6,t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & w_{7,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{8,t} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{9,t} \end{pmatrix}}_{\mathbf{S}} \times \underbrace{\begin{pmatrix} P_{y_1,t} \\ P_{y_2,t} \\ P_{y_3,t} \\ P_{y_4,t} \\ P_{y_5,t} \\ P_{y_6,t} \\ P_{y_7,t} \\ P_{y_8,t} \\ P_{y_9,t} \end{pmatrix}}_{\mathbf{z}_{K,t}}$$

[10] The original version of a hierarchical time series by Athanasopoulos et al. (2009) is a simple summation of base-level series. With weighted averages, such as the inflation rate, the index weights have to be incorporated into the summation matrix.

[11] This is a modification of the structure in Capistrán et al. (2010), that makes it aggregation consistent – a necessary property for HTS modelling.

This allows us to obtain forecasts of the unweighted series at each level of disaggregation as $\mathbf{z_{K,t}}$ contains the unweighted bottom-level series. The incorporation of the weights into the summation matrix keeps them exogenous from the forecast estimation of the inflation rates at the different levels. The forecasts in $\mathbf{Z}_t$ contain the inflation rate at the aggregate level, and the *weighted* series at the disaggregated levels – indicated by the superscript w. It can easily be checked that $\mathbf{Z}_t$ contains the actual contribution of the respective series to the aggregate inflation rate, making the hierarchical time series aggregation consistent:

$$\underbrace{\sum_{i=1}^{9} P^w_{y_i,t}}_{\text{Level 3}} = \underbrace{\sum_{i\in\{AA,AB,BA,BB\}} P^w_{y_i,t}}_{\text{Level 2}} = \underbrace{\sum_{i\in\{A,B\}} P^w_{y_i,t}}_{\text{Level 1}} = \underbrace{P_{Y,t}}_{\text{Level 0}}$$

Thus we can obtain valid forecasts of all series in the hierarchical time series, $\mathbf{Z}_t$, applying the different aggregation approaches for hierarchical time series (presented in the following section).

It is desirable to have forecasts of the unweighted series at the disaggregated levels – e.g., the forecast of industry level inflation, instead of the forecast of industry level inflation's contribution to aggregate inflation. This can be addressed easily: Noting that the h-step ahead forecasts of the quasi-exogenous index weights are $\hat{w}_{i,t+h|t} = w_{i,t}$, we can readily retrieve the forecasts for the unweighted series by dividing by the corresponding weights.

Aggregation Approaches. We consider the following aggregation approaches to forecasting the mean and variance of aggregate inflation, adapting the Hierarchical Time Series framework:

- *Top-down:* Forecast aggregate series and then disaggregate based on historical proportions (Gross and Sohl, 1990) or forecast proportions (Athanasopoulos et al., 2009). The methods can produce different forecasts for the disaggregated levels (Level 1 & Level 2), but all produce the same forecast for the aggregate level (Level 0). When the focus is on forecasting aggregate inflation rate/volatility, top-down methods are equivalent to an aggregate forecast.

- *Bottom-up:* Forecast disaggregate series at the lowest level and then aggregate up to forecast at higher levels. The argument in favour of the bottom-up approach is that the bottom-level series contain valuable information (e.g., different seasonal patterns in the bottom-level series). At the same time, smoothing noisy bottom-level series may lead to better forecasts of the aggregate. The empirical results are inconclusive. Top-down approaches do outperform bottom-up forecasts when the bottom-level data is noisy (e.g., Shlifer and Wolff, 1979; Fliedner, 1999; Hendry and Hubrich, 2006). There is also support the efficacy of bottom-up forecasts over top-down forecasts (e.g., Orcutt et al., 1968; Collins, 1976; Dunn et al., 1976; Dangerfield and Morris, 1992; Zellner and Tobias, 2000). While it seems intuitive that the bottom-up approach might provide higher bottom-level accuracy, and a top-down approach higher top-level accuracy, Hyndman et al. (2011) conclude that the bottom-up method performs significantly better than the conventional top-down method even for top-level forecasts.

Kahn (1998) suggests a hybrid approach, based on the argument that the efficacy of aggregation depends on the covariance structure of the constituent series (Tiao and Guttman, 1980; Kohn, 1982).

- *Middle-out*: A hybrid approach involving forecasting at intermediate levels, for aggregation to the higher levels (using a bottom-up approach), and disaggregation to the lower levels (using a top-down approach).

- *Optimal Combination*: Forecast each series in the hierarchy not heeding "aggregation consistency". Then optimally combine the forecasts to generate revised forecasts that are aggregation consistent and as close as possible to univariate forecasts. The reconciliation of forecasts is usually based on a Generalised Least Squares (GLS) estimator, but in practice reverts to OLS (Hyndman et al., 2011) or WLS (Hyndman et al., 2016a) due to the difficulty of estimating the covariance matrix of the reconciliation errors, which is non-identifiable (Wickramasuriya et al., 2015). Several adjustments to the general approach have been proposed: Di Fonzo and Marini (2011) and Hyndman et al. (2016a) exploit the sparsity of the linear system, thereby making it possible to reduce computational complexity when a very large number of time series is involved. van Erven and Cugliari (2015) propose the Game-Theoretically OPtimal (GTOP) method that guarantees that the total weighted quadratic loss of the reconciled forecasts will never be greater than the total weighted quadratic loss of the base forecasts. Wickramasuriya et al. (2015) propose the Minimum Trace (MinT) reconciliation that minimises the sum of variances of the reconciled forecast errors under the assumption of unbiasedness.

Forecasting Approaches. We assess the forecast accuracy of each of these four aggregation approaches, using six different univariate forecasting methods. Three of the methods are commonly used in hierarchical time series forecasting and are included in the R package *hts* (Hyndman et al., 2016b) – ARIMA (Box and Jenkins, 1970), ETS (introduced/extended by: Pegels, 1969; Gardner, 1985; Hyndman et al., 2002; Taylor, 2003), and a naïve Random Walk forecast as benchmark. We use the selection algorithms based on minimising AICc proposed by Hyndman and Khandakar (2008) to fit the ARIMA and ETS models. In addition, the following methods, which have found favour in forecasting competitions as accurate, robust, and reliable, are also employed:

- *Damped Trend* (Gardner and McKenzie, 1985): These models deal with the problem that exponential smoothing methods with constant trend tend to over-forecast, by introducing a parameter that dampens the trend to a flat line some time in the future. The superior performance of the damped trend model compared to a range of other methods is documented (Fildes and Ord, 2002; Armstrong, 2006). We employ an additive damped trend model with additive errors and an additive seasonal component – this corresponds to an $ETS(A, A_d, A)$ model in the general notation of (Hyndman et al., 2008).[12]
- *Theta method* (Assimakopoulos and Nikolopoulos, 2000): The Theta method has been found to produce the most accurate forecasts for monthly data in the M3 forecasting competition (Makridakis and Hibon, 2000) and come to serve as a benchmark in more recent forecasting competitions (Athanasopoulos et al., 2011). It is also relatively simple and computationally fast (Nikolopoulos et al., 2012). The Theta method is applied to deseasonalised time series (usually based on the multiplicative classical decomposition).[13] The forecasts obtained with the Theta method are equivalent to Simple Exponential Smoothing with drift (Hyndman and Billah, 2003).
- *Dynamic Optimised Theta Model* (Fiorucci et al., 2016b) is a generalization of the standard Theta method. DOTM produced more accurate forecasts than the standard Theta model for almost all combinations of type of data and frequency of the M3 time series (Fiorucci et al., 2016b) – the disadvantage is its higher computational intensity, which may be important when the number of base-level series is very large. With only 356 series (for the mean equation) and 15 series (for the variance equation), we have no difficulty in using DOTM.[14]

  Some special features of our data are to be noted. Thomakos and Nikolopoulos (2014) document that the Theta method performs especially well with trended series. Seasonality is a more dominant feature in our data than trend, and this might favour ARIMA and ETS models which incorporate seasonality in estimation. The Theta method only forecasts the deseasonalised series and then reseasonalises the data based on the multiplicative classical decomposition. The Theta method is inferior to other methods for forecasting monthly data with strong seasonality (Athanasopoulos et al., 2011).

[12] The damped trend model is part of the ETS framework. So the ETS algorithm will also select a damped trend specification for some of the series in the hierarchy. However, it is common to use it as a separate method – thereby forcing all series into a damped trend specification.

[13] Fiorucci et al. (2016b) argue that the seasonality test employed might not work well if the time series has one or more unit roots with a slow decay in the autocorrelation function. Since inflation rates are typically stationary, this is not an issue.

[14] The variants of the Theta models were estimated using the *forecTheta* package in R (Fiorucci et al., 2016a). The provided seasonality test was used in order to choose between additive and multiplicative decomposition. Model parameters were optimised using the Nelder-Mead algorithm.

The accuracy of the forecasts produced with these six forecasting methods is evaluated based on MAE[15]. Our forecast accuracy evaluation involves time series cross-validation based on training sets with a minimum of 180 observations, and constant length test sets. This involves separate analyses for 1 month, all the way up to 12 months horizons.[16]

Forecast Combination: Dynamic Model Switching. So far, we have been concerned with methods for determining the best individual forecast model, and aggregation procedure, for disaggregated inflation forecasting. In the next step, we take on board the fact that the models that produce most accurate forecasts often differ depending on different economic conditions. For example, Philips Curve-based models are more accurate during recessions, but do not consistently outperform the best univariate models at other times (Stock and Watson, 2008). We adopt a dynamic model selection (DMS) approach, which leverages the ability of different models to produce more accurate forecasts under different conditions. DMS is an extreme variant of the forecast combination method in which a combination weight of 1 is allocated to one of the candidate forecasting models, in a time-varying fashion depending on pre-defined criteria.

The literature on DMS is limited but promising: Belmonte and Koop (2013) use switching linear Gaussian state space models to forecast inflation; McMillan (2014) uses in-sample criteria in order to select between linear and nonlinear models for stock return forecasting; Buncic and Moretto (2015) use a dynamic model selection and averaging framework for copper price forecasts. Bagdatoglou et al. (2016) find that a dynamic model selection and averaging algorithm can improve forecasting accuracy for US inflation significantly compared to the UC-SV model of Stock and Watson (2008).

The forecast combination approach typically uses a linear combination of forecasts obtained from different models for the same time series, motivated by the view that all models of real world data generation processes are mis-specified, and combining forecasts across models can decrease model uncertainty and improve accuracy, by exploiting the different strengths of different models while compensating for their weaknesses.[17] One challenge of applying forecast combination to hierarchical time series is that the requirement of lack of high collinearity between the different forecast series is rarely satisfied. Similar methods – such as ETS and damped trend, or the Theta method – tend to produce collinear forecasts. The model switching approach that we employ does not break down under collinearity (McMillan, 2014). It can be seen as a boundary case of forecast combination.

We have a number of candidate models – 26 of them, resulting from combining every aggregation approach with every forecasting approach. It is likely that different models forecast better for different parts of the sample, differ from each other. Choosing a single model out of many for the entire sample is likely to lead to poor out-of-sample forecasts. We noted earlier that the inflation rate series can be considered to have two volatility regimes, high and low. It is well-established that inflation rate and inflation volatility are strongly interconnected (Friedman-Ball hypothesis & Cukierman-Melzer hypothesis). The model switching approach allows us to use in-sample variance as criterion to select the best model for each point in time. The

[15] Other standard measures such as RMSE, MPE, MAPE, MASE can be readily used. Since all the forecasts are computed for a single series – aggregate inflation rate, or aggregate inflation variance – we prefer MAE for its easy interpretation, because it is less sensitive to outliers than MSE or RMSE and it avoids common problems with MPE and MAPE in the presence of zero or very small values in some series. It should be noted that MAE is a scale-dependent measure and could not be used in a comparison between series (Hyndman and Koehler, 2006).

[16] The suitability of cross-validation for accuracy assessment with time series data is discussed in Bergmeir et al. (2015).

[17] Different methods have been proposed and implemented: Clemen (1989) argues that simple averaging of all available forecasts is a successful and robust method; Armstrong (2001) suggests the use of trimmed means to avoid sensitivity to extreme values. Stock and Watson (2004) find that symmetric 5 % trimming performed about the same as simple averaging, while Jose and Winkler (2008) find that trimming of 10 - 30 % or Winsorizing of 15 - 45 % can lead to improved accuracy compared to simple averages using data from the M3 forecasting competition; Granger and Ramanathan (1984) calculate the combination weights using an OLS regression; Aiolfi and Timmermann (2006) group forecasts into several clusters using a k-means algorithm, final forecasts are then obtained by averaging forecasts of the historically better performing cluster; and Hsiao and Wan (2014) introduce several ways of eigenvector-based forecast combination.

interconnection between inflation rate and volatility makes it more likely that changes in in-sample variance affect the underlying data generating process. This is the fundamental rationale for switching between models.[18] The model switching approach involves recursive estimation and in-sample forecast evaluation to guide the switching between forecasts produced using different models.

We adopt the following model switching rule (for both the mean and variance forecasts): Split up the validation set, $\mathbb{S}^{CV}$ into two subsets, a high-volatility set, $\mathbb{S}^{CV}_{high}(x)$ and a low-volatility set, $\mathbb{S}^{CV}_{low}(x)$. The elements of the high-volatility subset relate to those time points where the unconditional inflation volatility (as measured by the variance of the aggregate inflation rate over the past 6 months) exceeds the x^{th} quantile of in-sample variance distribution, $\Phi(\sigma^2_{sample}, x)$:[19]

$$\mathbb{S}^{CV}_{high}(x) = \left\{y_t \in \mathbb{S}^{CV} | \sigma^2_t \leq \Phi(\sigma^2_{sample}, x)\right\}$$
$$\mathbb{S}^{CV}_{low}(x) = \left\{y_t \in \mathbb{S}^{CV} | \sigma^2_t > \Phi(\sigma^2_{sample}, x)\right\}$$

This leaves the choice of x – the threshold that optimally splits the support of the sample variance distribution, such that validation set is split into the two sub-samples in a way that allows us to minimise the MAE of the final forecast.

We adopt the following procedure:

1. Compute $\mathbb{S}^{CV}_{high}(x)$ and $\mathbb{S}^{CV}_{low}(x)$ for each possible x, i.e. for each percentile of the sample variance distribution,

2. For each value of x (100 cases), produce forecasts for $\mathbb{S}^{CV}_{high}(x)$ with all 26 candidate models and select the model with the highest accuracy (lowest MAE) for the given high-volatility subset – denote this optimal model for the high-volatility subset for a given x by $model^{opt}_{high}(x)$; for each x, produce forecasts for $\mathbb{S}^{CV}_{low}(x)$ with all 26 candidate models and select the model that shows the highest accuracy (lowest MAE) for the given low-volatility subset – denote this optimal model for the low-volatility subset for a given x by $model^{opt}_{low}(x)$

3. For each x, use $model^{opt}_{high}(x)$ to produce h-step ahead forecasts, $\hat{y}^{high}_{t+h|t}$, if $y_t \in \mathbb{S}^{CV}_{high}(x)$; and $model^{opt}_{low}(x)$ to produce h-step ahead forecasts, $\hat{y}^{low}_{t+h|t}$, if $y_t \in \mathbb{S}^{CV}_{low}(x)$. This yields forecasts for the entire cross-validation set, by combining the forecasts from the two subsets:

$$\hat{y}_{t+h|t}(x) = \begin{cases} \hat{y}^{high}_{t+h|t}, & \text{if } y_t \in \mathbb{S}^{CV}_{high}(x) \\ \hat{y}^{low}_{t+h|t}, & \text{if } y_t \in \mathbb{S}^{CV}_{low}(x) \end{cases}$$

4. Compute MAE for the forecasts of the cross-validation set for each x, and select the MAE-minimizing x, x_{opt}, as a splitting point for our switching rule, which now becomes:

$$\hat{y}_{t+h|t}(x_{opt}) = \begin{cases} \hat{y}^{high}_{t+h|t}, & \text{if } y_t \in \mathbb{S}^{CV}_{high}(x_{opt}) \\ \hat{y}^{low}_{t+h|t}, & \text{if } y_t \in \mathbb{S}^{CV}_{low}(x_{opt}) \end{cases}$$

To summarise, the general approach is a data-driven procedure to select the forecast model depending on an in-sample criterion as it varies over time – in this study, the unconditional sample variance corresponding to the point in time at which the forecast is made. Due to the greater flexibility relative to a fixed forecast model, this approach has the potential to improve forecast accuracy.

[18] McMillan (2014) used the best-fitting model of the previous period (according to AIC) as in-sample criterion to select a forecast model in a recursive framework.

[19] The switching rule is presented for the inflation rate for illustration, but applies to inflation volatility forecasts analogously.

Forecasting System: Inflation Mean & Variance. The empirical strategy is to estimate time-varying expectation of the inflation rate, fitting the model (e.g., ARIMA, ETS) to Y_t, and to use the fitted conditional mean $\hat{Y}_t$ as the expected inflation rate in the variance equation:[20]

$$\hat{\sigma}^2_{Y_t} = \lambda\sigma^2_{t-1} + (1-\lambda)Y^2_{t-1} - 2(1-\lambda)Y_{t-1}\hat{Y}_t + (1-\lambda)\hat{Y}^2_t$$

The above equation illustrates how the fitted conditional mean enters the variance equation through the EWMA calculation. In order to keep the presentation simple we show this only for the aggregate inflation; however, this system that comprises of a mean model and a variance model can easily be applied at various levels of disaggregation (e.g., $\hat{y}_{it}$ entering into the EWMA calculation for $\hat{\sigma}^2_{y_{it}}$ at the product level).

To obtain the mean forecast, we use the subcomponents of the product-level inflation rates as input. These base-level series are combined with the index weights into a hierarchical time series. Forecasts are produced for the aggregate mean inflation rate using the aggregation approaches and forecast methods described above.

The mean model fit will be used for the calculation of EWMA-smoothed base-level (co)variances (and, consequently, to create the input series of the variance model). Before this, the optimal EWMA decay parameter, λ, must be determined. Our strategy, consistent with inflation volatility literature, is to use the conditional volatility as estimated by a GARCH(1,1) as proxy for the 'actual' values of volatility, and to estimate EWMA-smoothed aggregate volatility using a large range of potential decay parameter values. The RMSE-minimizing λ can then be selected and – together with the fitted values of the base-level parts from the best mean model – used for the computation of the EWMA-smoothed base-level covariance matrix. By applying the index weights to the corresponding terms of the smoothed base-level covariance matrix, we directly compute the parts of variance component and the covariance component as inputs for the HTS volatility forecasts that are produced using the forecasting and aggregation methods described above.

5. Results

We now present an evaluation of forecast accuracy of the hierarchical time series based mean and volatility forecasts. With the exception of the Theta method variants, the models were estimated using data with no prior seasonal adjustment, extracting seasonality in the estimation.[21] For the Theta method models, the series were deseasonalised first, using the algorithm provided in Fiorucci et al. (2016a) which bases the choice between additive and multiplicative classical decomposition on a seasonality test, and were reseasonalised after the estimation. The tables in this section present the relative accuracy of the forecasts, i.e., $\frac{MAE_{model}}{MAE_{opt}}$ where MAE_{opt} is the accuracy of the best model.

Mean Model – Results. The hierarchical time series data that is used to estimate the mean models has 4 levels of disaggregation (from lowest to highest): Level 0 – is the aggregate inflation rate, Level 1 – is

[20] An alternative would be to think of the expected inflation rate as a linear combination of the mean model, $\hat{Y}_t$, and the target inflation rate, μ:

$$\gamma\hat{Y}_t + (1-\gamma)\mu$$

The EWMA variance equation is then:

$$\begin{aligned}\hat{\sigma}^2_{Y_t} =& \lambda\sigma^2_{t-1} + (1-\lambda)Y^2_{t-1} - 2(1-\lambda)\gamma Y_{t-1}\hat{Y}_t \\ & - 2(1-\lambda)(1-\gamma)Y_{t-1}\mu + (1-\lambda)\gamma\hat{Y}^2_t \\ & + (1-\lambda)(1-\gamma)\mu^2\end{aligned}$$

[21] Alternatively, deseasonalised base-level series could be used. However, since the forecasting frameworks – ARIMA and Exponential Time Series Smoothing (including Damped Trend) – are able to incorporate seasonality in the model, we choose to use seasonal base-level series to avoid losing valuable base-level information.

the 15 industry-level inflation rates (e.g., Food, Housing, etc.), Level 2 – is the 85 product-level inflation rates (e.g., Bread, Furniture, Pet Care, etc.), and Level 3 – decomposes each product-level inflation rate into common, industry, and idiosyncratic parts. Consequently, the Bottom-Up approach works with the 255 base-level series and, using the summation matrix, aggregates the fitted values to obtain a fit for aggregate inflation rate; the Middle-Out (Level 2) approach fits the product-level series and aggregates to the aggregate inflation rate[22]; the Middle-Out (Level 1) approach fits the industry-level series and aggregates to the aggregate inflation rate; and the Top-Down approach works directly with the aggregate inflation rate series. The Optimal Combination approach fits all 356 constituent series of the hierarchical time series, at their different levels of disaggregation, and then applies the OLS method of Hyndman et al. (2011) to reconcile the forecasts. Table 1 presents relative forecast accuracy compared to the best model of the respective horizon for the full validation set – the absolute MAE value is presented in brackets for the best model for each horizon.

Table 1: Cross-Validated Test Set Accuracy - Mean Model.

	1m	2m	3m	4m	5m	6m	7m	8m	9m	10m	11m	12m
Naïve												
Naïve	1.915	1.860	1.925	1.874	1.879	1.846	1.868	1.884	1.899	1.893	1.897	1.854
Top-Down												
ETS	1.080	1.116	1.170	1.178	1.202	1.229	1.246	1.262	1.263	1.257	1.254	1.257
ARIMA	1.172	1.160	1.163	1.167	1.174	1.186	1.197	1.206	1.207	1.208	1.209	1.212
Theta	1.093	1.117	1.151	1.165	1.190	1.210	1.224	1.234	1.232	1.229	1.227	1.231
Damped Trend	1.106	1.128	1.183	1.200	1.230	1.259	1.278	1.292	1.293	1.287	1.282	1.282
Dynamic Optimised Theta	1.093	1.115	1.147	1.160	1.184	1.203	1.217	1.226	1.224	1.221	1.220	1.223
Middle Out (Level 1)												
ETS	1.008	1.021	1.048	1.072	1.088	1.106	1.119	1.126	1.128	1.129	1.132	1.137
ARIMA	1.017	1.001	1.011	1.023	1.028	1.030	1.031	1.029	1.025	1.021	1.018	1.018
Theta	1.047	1.059	1.094	1.113	1.136	1.154	1.166	1.174	1.175	1.176	1.176	1.183
Damped Trend	1.014	1.019	1.060	1.089	1.115	1.140	1.157	1.169	1.172	1.174	1.178	1.183
Dynamic Optimised Theta	1.043	1.055	1.090	1.109	1.132	1.150	1.163	1.170	1.171	1.172	1.173	1.179
Middle Out (Level 2)												
ETS	1.018	1.003	1.027	1.039	1.048	1.062	1.073	1.079	1.080	1.082	1.084	1.087
ARIMA	1.021	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		(0.0023)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Theta	1.085	1.069	1.084	1.092	1.099	1.116	1.125	1.132	1.132	1.132	1.139	1.149
Damped Trend	1.065	1.045	1.078	1.104	1.121	1.137	1.148	1.157	1.160	1.160	1.165	1.171
Dynamic Optimised Theta	1.110	1.084	1.090	1.096	1.104	1.121	1.131	1.138	1.137	1.137	1.144	1.154
Bottom Up												
ETS	1.088	1.114	1.169	1.201	1.230	1.250	1.258	1.269	1.265	1.261	1.263	1.262
ARIMA	1.092	1.076	1.079	1.083	1.088	1.088	1.089	1.085	1.080	1.077	1.072	1.069
Theta	1.154	1.166	1.204	1.218	1.238	1.259	1.272	1.282	1.284	1.284	1.284	1.288
Damped Trend	1.083	1.101	1.159	1.185	1.212	1.244	1.261	1.278	1.285	1.284	1.287	1.290
Dynamic Optimised Theta	1.144	1.157	1.193	1.207	1.227	1.247	1.260	1.268	1.269	1.270	1.270	1.274
Optimal Combination												
ETS	**1.000**	1.022	1.070	1.089	1.113	1.138	1.155	1.169	1.170	1.168	1.170	1.175
	(0.0022)											
ARIMA	1.061	1.053	1.060	1.068	1.076	1.086	1.094	1.098	1.098	1.099	1.099	1.101
Theta	1.087	1.109	1.144	1.158	1.182	1.202	1.216	1.226	1.225	1.222	1.220	1.224
Damped Trend	1.089	1.110	1.164	1.182	1.213	1.242	1.261	1.276	1.277	1.272	1.268	1.269
Dynamic Optimised Theta	1.087	1.108	1.140	1.153	1.176	1.196	1.209	1.219	1.217	1.215	1.213	1.217

Accuracy assessment based on time series cross-validation provides clear support for the disaggregated approach. The best disaggregated model beats the best top-down univariate approach for all horizons, by a range between 8 % (for a one-step ahead forecast) and 21 % (for a twelve-step ahead forecast). For all horizons apart from 1 month, the Middle Out (Level 2) ARIMA model produces the most accurate forecasts, i.e., the best approach to forecasting the conditional mean of aggregate inflation is to forecast product-level inflation rates (Level 2) using (seasonal) ARIMA models and then aggregate these forecasts.

[22]Middle-Out approaches apply a bottom-up approach to obtain estimates for levels of lower disaggregation and a top-down approach to obtain estimates for levels of higher disaggregation.

These results are consistent with previous empirical findings on both aggregation approaches and forecasting methods: Our conclusions are similar to those of Athanasopoulos et al. (2011) about the dominance of ARIMA and ETS techniques compared to Naïve, Damped Trend, and Theta method for non-trended data that has strong seasonality. The results also support findings of a variety of studies evaluating the potential of disaggregated methods (e.g., Kahn, 1998) that conclude that bottom-up methods do not usually produce good aggregate forecasts: We find that the hybrid approaches (Middle-Out and Optimal Combination) outperform bottom-up; however, it should be noted that the best bottom-up approach outperforms top-down for all horizons except for 1-month forecasts – this suggests that the value of disaggregated information (in this case, mainly the different seasonal patterns of the bottom-level series) outweighs the cost of noisy bottom-level data. Finally, while both Theta method variants do not seem well-suited for forecasting seasonal, non-trended data, the dynamic optimised Theta method produces slightly better results than the original Theta method for 4 out of 5 aggregation approaches. All methods produce much better results than a naïve approach.

In order to present the information from the table in a concise way, Fig. 6 plots the mean ranking, as well as the range of rankings, for each method out of the 26 models. This shows much better that (a) ARIMA and ETS methods are dominant for seasonal, non-trended monthly data, and (b) how valuable the disaggregated modelling approach is for inflation forecasting, given that the best aggregate model (Top Down ARIMA) has an average ranking of 16.5 out of 26 models.

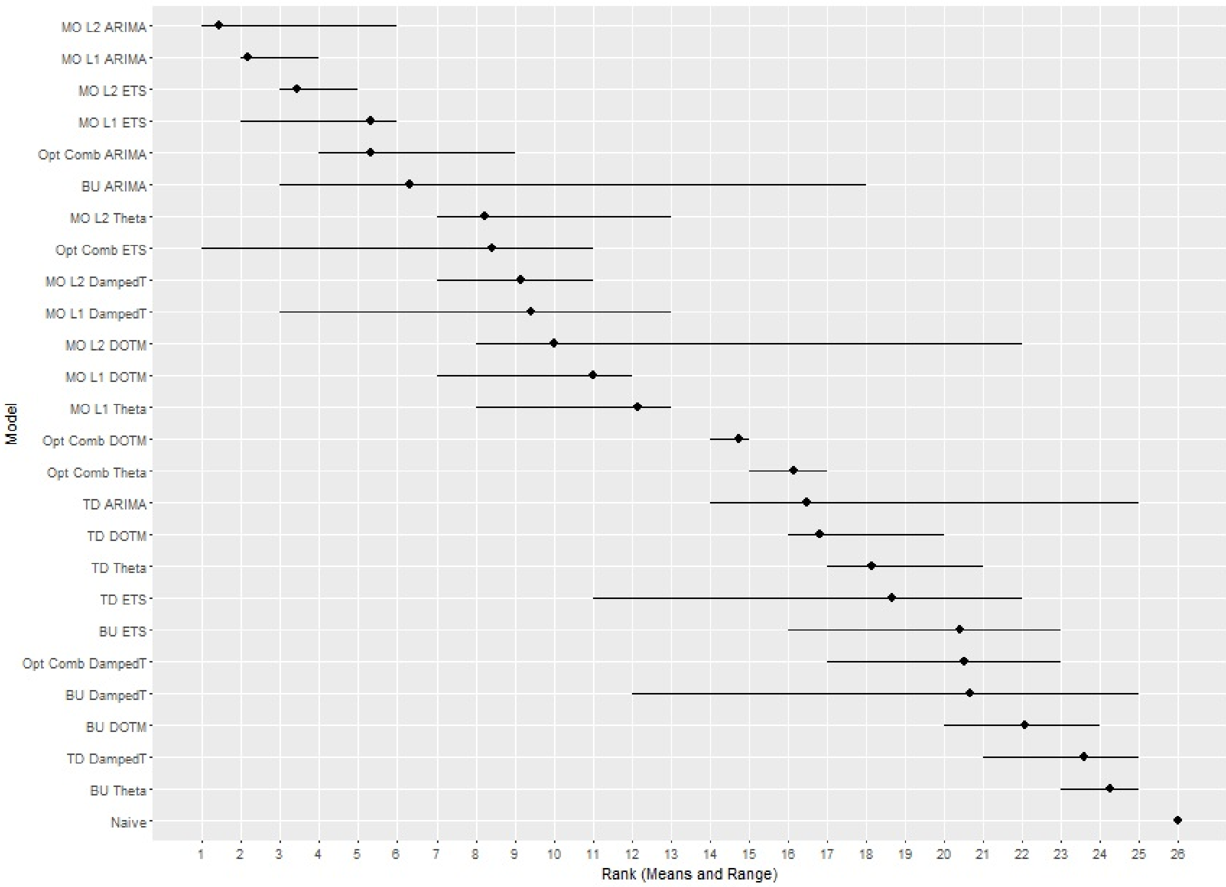

Fig. 6. Ranking (Mean and Range) of the Mean Model Approaches.

Table 2 shows the test statistics and p-values of the Diebold-Mariano test (with a quadratic loss function) for predictive accuracy, comparing the performance of the best disaggregated model with the best aggregate model. For all horizons except for 1-month forecasts, the disaggregated model forecasts significantly better, at least at the 10 % level. The disaggregated models are seen to be relatively more valuable for multi-horizon forecasts.

Table 2: Diebold-Mariano Tests of Predictive Accuracy

	1m	2m	3m	4m	5m	6m
Best Aggregate	Opt Comb ETS	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA
Best Disagg.	TD ETS	TD DOTM	TD DOTM	TD DOTM	TD ARIMA	TD ARIMA
DM Statistic	-0.80	-1.62	-1.57	-1.41	-2.12	-2.19
p-value	0.210	0.054	0.059	0.080	0.018	0.015
	7m	8m	9m	10m	11m	12m
Best Aggregate	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA
Best Disagg.	TD ARIMA	TD ARIMA	TD ARIMA	TD ARIMA	TD ARIMA	TD ARIMA
DM Statistic	-2.28	-2.31	-2.20	-2.13	-2.04	-2.04
p-value	0.012	0.011	0.015	0.018	0.022	0.022

Mean Model – Dynamic Model Switching. In order to assess the value of allowing for different models in low-volatility and high-volatility phases, we applied the dynamic model switching rule outlined in Section 4. Table 3 presents the optimal (MAE minimizing) quantiles ($\Phi(\sigma^2_{sample}, x_{opt})$) of the sample variance distribution that divide the validation set into the high-volatility and low-volatility subsets (denoted as 'Split Variance'), and the best models selected for the two subsets.

Table 3: Switching Rule Mean Model: Results for Different Horizons.

	1m	2m	3m	4m	5m	6m
Best High Vol	MO L1 ARIMA	MO L1 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA
Best Low Vol	BU ETS	BU ETS	BU ETS	BU ETS	BU DampedT	BU Theta
Split Variance	5.45e-06	5.42e-06	5.29e-06	5.29e-06	5.29e-06	5.29e-06
	7m	8m	9m	10m	11m	12m
Best High Vol	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L2 ARIMA	MO L1 ARIMA	MO L1 ARIMA
Best Low Vol	MO L2 Theta	MO L2 Theta	MO L2 ETS	MO L2 ETS	MO L2 ETS	MO L2 ETS
Split Variance	5.29e-06	5.29e-06	8.99e-06	8.99e-06	1.79e-05	2.29e-05

The application of the switching rule delivers some interesting insights in the previous findings: We find that the dominance of the Middle Out ARIMA approaches stems from their better performance at times of higher volatility. In phases of lower volatility, Bottom-Up ETS, Damped Trend, and Theta models (all of which belong to the exponential smoothing family) are best up to 6-months, and Middle Out (Level 2) Theta and ETS are best for longer horizons. It makes sense that bottom-up models are well-suited for forecasting at times of low volatility, as they contain valuable additional information and noise is often considered to be positively correlated with volatility (e.g., Bandi and Russell, 2006).

Fig. 7 shows the MAEs of the switching-rule models in Table 3, comparing their accuracy to the best single-method HTS model. The switching-rule forecasts are between 1.8 % and 9.3 % more accurate compared to the best single-method HTS model, depending on horizon. For all horizons, the Diebold-Mariano test leads to the conclusion that the switching rule forecast significantly outperforms the best single-method HTS model, confirming the potential of the technique.

Variance Model – Results. The fitted values of the mean model are used in the EWMA calculation of the base-level covariance matrix that is subsequently used to compute the parts of the VC and the CC – the inputs of the variance model. The decay parameter for EWMA, λ, was selected by the RMSE minimization procedure described above, which returned a λ of 0.83 (Fig. 8).

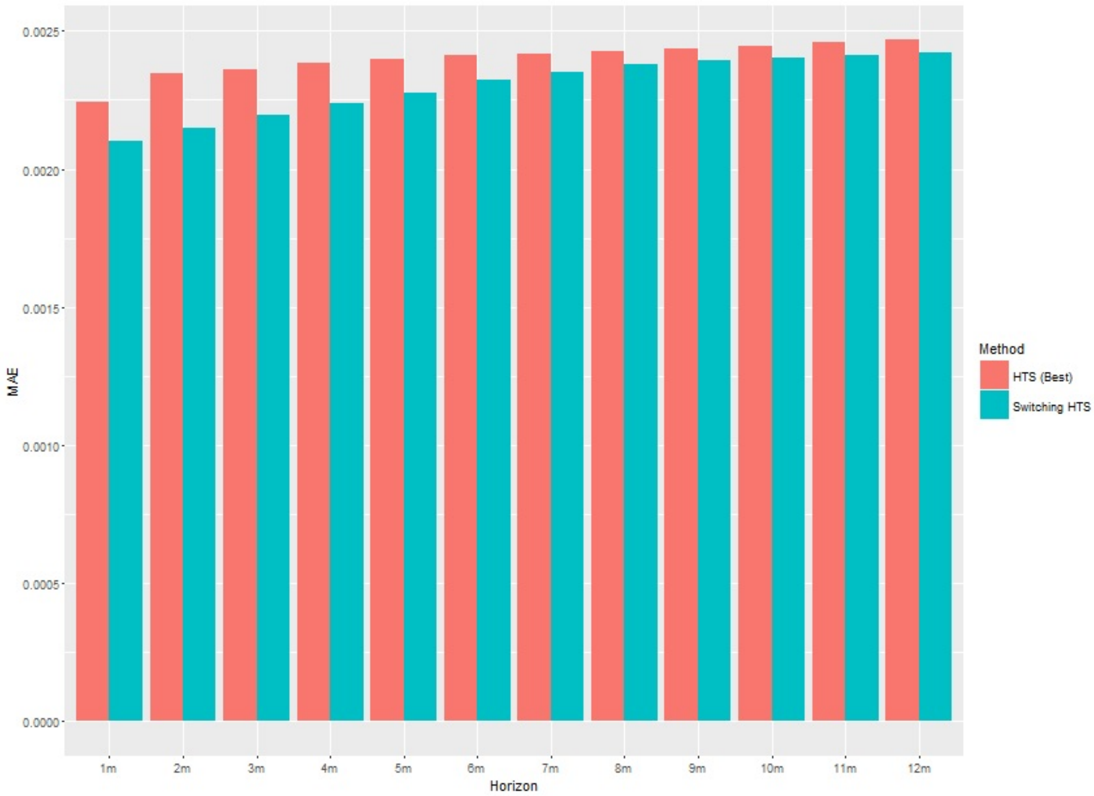

Fig. 7. Mean Model: MAEs of Switching-Rule vs. Best Single-Method HTS.

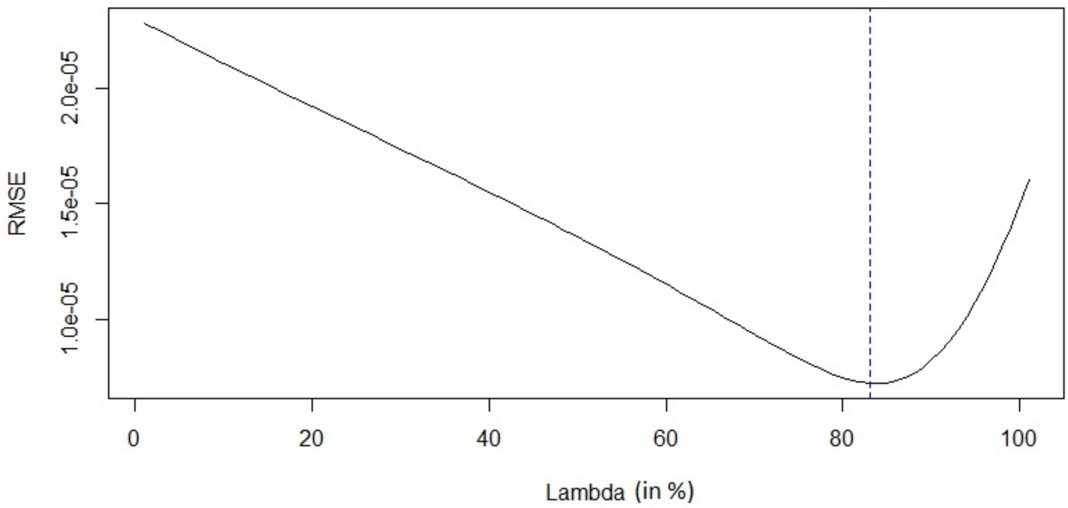

Fig. 8. RMSE-Optimization of EWMA persistence parameter λ.

The hierarchical time series used as input for the variance models has 3 disaggregation levels (from lowest to highest): Level 0 – aggregate inflation variance, Level 1 – variance component (VC) and covariance component (CC), Level 2 – the 6 subparts each of VC and CC (Section 2.2).

Table 4: Cross-Validated Test Set Accuracy - Variance Model.

	1m	2m	3m	4m	5m	6m	7m	8m	9m	10m	11m	12m
Naïve												
Naïve	1.008	1.006	1.004	1.004	1.003	1.003	1.002	1.006	1.003	1.003	1.003	1.003
Top-Down												
ETS	1.350	1.365	1.388	1.519	1.621	1.694	1.783	1.872	1.993	2.206	2.504	2.899
ARIMA	1.306	1.386	1.417	1.427	1.426	1.422	1.407	1.402	1.396	1.410	1.434	1.459
Theta	1.075	1.065	1.054	1.048	1.041	1.036	1.032	1.034	1.030	1.031	1.033	1.034
Damped Trend	1.319	1.252	1.198	1.164	1.131	1.118	1.102	1.104	1.103	1.114	1.120	1.124
Dynamic Optimised Theta	1.071	1.062	1.052	1.046	1.039	1.034	1.030	1.031	1.027	1.028	1.029	1.029
Middle-Out												
ETS	1.244	1.241	1.248	1.265	1.272	1.278	1.283	1.299	1.310	1.333	1.358	1.385
ARIMA	1.108	1.038	1.029	1.020	1.019	1.004	1.000	**1.000**	1.001	1.012	1.025	1.036
								(4.21e-06)				
Theta	1.064	1.058	1.049	1.044	1.038	1.033	1.030	1.032	1.028	1.029	1.031	1.032
Damped Trend	1.231	1.195	1.174	1.166	1.157	1.158	1.157	1.161	1.163	1.179	1.193	1.205
Dynamic Optimised Theta	1.061	1.055	1.047	1.042	1.036	1.031	1.028	1.029	1.025	1.026	1.026	1.027
Bottom-Up												
ETS	1.273	1.209	1.116	1.071	1.029	1.004	1.002	1.017	1.025	1.033	1.033	1.031
ARIMA	1.141	1.254	1.257	1.240	1.235	1.212	1.194	1.187	1.182	1.192	1.206	1.225
Theta	1.002	1.002	1.001	1.001	1.001	1.001	1.001	1.006	1.002	1.003	1.003	1.004
Damped Trend	1.728	1.665	1.594	1.566	1.535	1.514	1.518	1.535	1.545	1.558	1.572	1.580
Dynamic Optimised Theta	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.004	**1.000**	**1.000**	**1.000**	**1.000**
	(1.84e-06)	(2.31e-06)	(2.77e-06)	(3.14e-06)	(3.46e-06)	(3.77e-06)	(4.03e-06)		(4.37e-06)	(4.43e-06)	(4.47e-06)	(4.50e-06)
Optimal Combination												
ETS	1.235	1.244	1.247	1.297	1.326	1.348	1.382	1.424	1.475	1.564	1.685	1.844
ARIMA	1.183	1.172	1.159	1.155	1.156	1.155	1.153	1.156	1.158	1.174	1.195	1.216
Theta	1.065	1.059	1.049	1.044	1.038	1.033	1.030	1.032	1.028	1.029	1.030	1.032
Damped Trend	1.194	1.154	1.118	1.107	1.095	1.097	1.095	1.103	1.106	1.122	1.133	1.140
Dynamic Optimised Theta	1.061	1.055	1.047	1.042	1.036	1.031	1.027	1.029	1.025	1.026	1.026	1.027

Turning to the variance forecast results (Table 4), again based on a time series cross-validation, several aspects are noteworthy:

- The value of the disaggregated approach is understated in the variance forecasts. It should be noted that disaggregated models by far outperformed the aggregate models for the conditional mean of inflation and that these estimates were used as inflation expectation in the computation of the EWMA-smoothed bottom-level covariance matrix. Using inflation expectations of higher accuracy (compared to ones from aggregate models) improves the *input data* that the variance forecasts are based on – this value is not incorporated in the forecast accuracy tables.

- For the variance forecast, the naïve model performs extremely well. This could be due to a number of different reasons:

  – Even though we used EWMA to compute the rolling-window covariances between series (rather than the equally-weighted SMA), the resulting smoothing of the input data removes some of the structure that might make more sophisticated models useful.

  – Over most of the observation period, the variance series is relatively stable at a low level – conditions that favour a naïve forecast, even more so given the lack of pronounced seasonality in the aggregate variance or its parts at the disaggregated levels. The only period that is characterised by high variance is the Great Recession, which caused a sudden spike in volatility that none of the models managed to capture very well;

  – The first part of this paper has established that since inflation targeting was introduced, VC (the part of aggregate variance due to bottom-level variances) and CC (the part of aggregate variance due to co-movement of the bottom-level units) are highly correlated (Correlation Coefficient = 84.94 %). This suggests that a middle-out approach cannot be expected to be of very high value for the variance forecast, as the 2 series do not add a lot of information compared to a forecast at the top level.

- The only models that can consistently beat the naïve forecast for all horizons are the Bottom-Up Theta model (except for 11 and 12-month forecasts) and the Bottom-Up Dynamic Optimised Theta model.

This can be rationalised: First, the variance series are not dominated by seasonality as was the case for inflation rate itself. These are favourable conditions for Theta models. Second, while the Theta model – as special case of simple exponential smoothing with drift – is also a relatively simple forecasting model, it is designed to model the local curvature of the data (through the second Theta line). This can explain why these methods can actually beat the naïve forecast, which does not separately model long term and local trends. The Dynamic Optimised Theta model is designed to optimise the line for the local curvature more flexibly and it is unsurprising that the model outperforms (slightly) the original Theta model for all horizons. Third, the first part of this paper has shown that on the bottom level (i.e., among the subparts of VC and CC), there is some variation – while the VC is dominantly driven by the variance of the idiosyncratic part and to some extent the variance of the industry part (two series again follow similar patterns), the CC is driven by the variance of the common part and the covariances of the idiosyncratic parts – two series that follow very different patterns. This variation on the bottom-level, which was identified through the two-stage decomposition, is information that improves forecast accuracy, even if only slightly.[23]

- The good performance of the Middle Out ARIMA (which is among the best 4 models for almost all horizons) seems like a curiosity at a first glance – ARIMA models do very poorly for all other aggregation approaches and the high correlation between the two series at the middle level also provides no explanation why this approach does relatively well. A more detailed analysis shows that this is due to Middle Out ARIMA being the best model in periods of high volatility (as presented in the switching model results). While we are able to identify its good performance at times of high volatility as reason for the good overall result, we have found no convincing explanation why the Middle Out ARIMA approach does so well in high-volatility phases – looking at the time points when volatility was above is 90 % quantile, we found that during these times, correlation between VC and CC were even higher with 92.69 %, which makes it very difficult to explain why this approach produces 30-40 % more accurate forecasts than the top-down ARIMA model.

Variance Model – Dynamic Model Switching. Table 5 presents the results of the application of the switching rule.

Table 5: Switching Rule Variance Model: Results for Different Horizons.

	1m	2m	3m	4m	5m	6m
Best High Vol	Opt Comb DampedT	Opt Comb DampedT	Opt Comb DampedT	Opt Comb DampedT	Middle Out ARIMA	Middle Out ARIMA
Best Low Vol	BU DOTM	BU DOTM	Naive	Naive	Naive	Naive
Split Variance	1.64e-05	1.68e-05	2.78e-05	1.68e-05	2.78e-05	2.57e-05

	7m	8m	9m	10m	11m	12m
Best High Vol	Middle Out ARIMA	Middle Out ARIMA	Middle Out ARIMA	Middle Out ARIMA	Middle Out ARIMA	Middle Out ARIMA
Best Low Vol	Naive	BU ETS	Naive	Naive	Naive	Naive
Split Variance	2.78e-05	1.60e-05	2.78e-05	2.78e-05	2.78e-05	2.78e-05

While overall the results coincide with the findings from the mean model section – simple models, such as naïve and Theta method tend to be best for low-volatility phases and more structured models, such as ARIMA are better suited for high-volatility phases – there are two notable features that are worth discussing: The Bottom-Up Dynamic Optimised Theta is only selected as the best model for short-term low-volatility forecasts – given that it is overall the best model for all horizons apart from one, this suggests that while not being much worse than a naïve model for longer-horizon forecasts in low-volatility phases, it produces more accurate forecasts than a naïve model in high-volatility phases; this can be ascribed to its ability to capture the local curvature of the series.

[23] Diebold-Mariano tests for the best disaggregated model compared to a naïve forecast returned p-values between 0.037 and 0.269 for the different horizons and therefore document that the forecast performance of the disaggregated models is better than the naïve forecast, but only for two horizons this improvement is significant at the 10 % level.

Another interesting aspect is that a model from the exponential smoothing family – Optimal Combination Damped Trend – is selected for short-horizon forecasts in high-volatility periods. However, this also shows the appeal of this switching-rule approach to some extent: Over the entire test set, the Optimal Combination Damped Trend model only ranks between 11 and 13 of the 21 candidate models even for 1- to 4-month forecasts due to its below-average forecasting performance in periods of low volatility – consequently, without the switching rule, we probably would not have identified this model's good ability to produce short-term forecasts in high-volatility phases. Fig. 9 plots the improvement in MAE of the switching-rule models compared to the best single-method HTS model for each horizon – the switching rule improves accuracy by between 1.8 % and 6.3 %, depending on horizon.

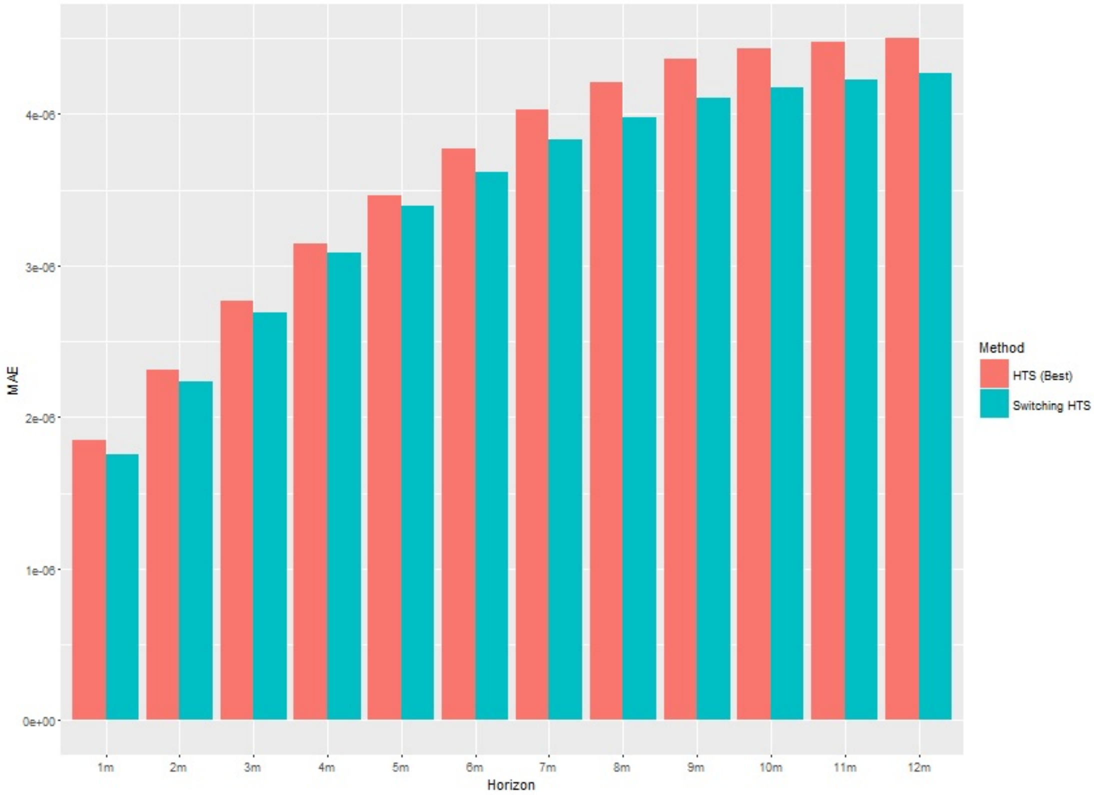

Fig. 9. Variance Model: Switching Rule vs. Best Single-Method HTS.

Validation of HTS Variance Models. For the purpose of validating the HTS variance models, we use GARCH-type models as benchmark. Estimating inflation uncertainty with GARCH-type models has a long history dating back to the seminal paper of Engle (1982) who first introduced the autoregressive conditional heteroskedasticity (ARCH) model, as well as Bollerslev (1986) who introduced the generalised ARCH model.

Using automated ARIMA selection, we find that a $SARIMA(1,0,1)(1,0,1)_{12}$ model fits best to describe the conditional mean. The ARCH-LM test rejected the null hypothesis of no ARCH effects (p-value: 0.25 %). ARCH/GARCH models are required.

Residual diagnostics revealed that the histogram of the residuals has a negative skew (Skewness: -0.896) and is leptokurtic (Kurtosis: 7.712). Previous research in the field documents that under these circumstances, the fit of GARCH-type models can be improved by using a leptokurtic error distribution – most commonly Student's t (proposed by Bollerslev, 1987) or GED (proposed by Nelson, 1991). Since the non-normality is primarily due to the largest outlier (a data point during the Great Recession), we also analysed normality without this outlier and find that without this data point, the histogram of the residuals approximates normality - the Jarque-Bera test statistic is 1.626 (p-value: 44.34 %). Given these results, we decided to estimate both GARCH-type models with Gaussian error distribution and ones with leptokurtic error distributions.

AIC-based selection returned a $SARIMA(1,0,1)(1,0,1)_{12} - GARCH(1,1)$ model with Gaussian error distribution as best fit for the inflation data (Table 6). We tried fitting asymmetric specifications (EGARCH, TGARCH), as well as Component GARCH due to its popularity in empirical inflation research, but none of the specifications were able to improve AIC compared to the GARCH(1,1). We also estimated the in-mean variants of all the mentioned models, but the in-mean effects were not significant. The AIC values of the estimated models are presented in Appendix A.

Table 6: Selected GARCH specification: $SARIMA(1,0,1)(1,0,1)_{12} - GARCH(1,1)$.

Variable	Coefficient	Std.Error	z-Statistic	p-value
AR(1)	0.657	0.182	3.615	0.0003
SAR(1)	0.983	0.008	125.442	0.0000
MA(1)	-0.420	0.226	-1.858	0.0632
SMA(1)	-0.887	0.027	-33.330	0.0000
Variance Equation				
C	1.60e-06	7.21e-07	2.221	0.0264
α (ARCH)	0.302	0.075	4.004	0.0001
β (GARCH)	0.395	0.175	2.261	0.0237

Having found the optimal GARCH-type model for our data, we turned to the validation of our models comparing the best HTS variance model against the selected SARIMA-GARCH model. Thus, to test the quality of our disaggregated forecasting system for inflation mean and variance, we again apply time series cross-validation to compare the best disaggregated model's accuracy with the SARIMA-GARCH model. Given the previous optimisation of the decay parameter, we use the conditional variance from an EWMA model fitted to the full aggregate data as proxy for 'actual' inflation variance.

Table 7: Forecasting Performance: Single-Method HTS vs. Switching Rule vs. SARIMA-GARCH.

	Best Single HTS	Best Switching HTS	SARIMA-GARCH	Δ
1m	1.845e-06	1.754e-06	4.470e-06	2.55
3m	2.768e-06	2.687e-06	4.333e-06	1.61
6m	3.772e-06	3.615e-06	4.551e-06	1.26
9m	4.366e-06	4.104e-06	4.776e-06	1.16
12m	4.500e-06	4.271e-06	4.642e-06	1.09

Table 7 compares the MAE accuracy of the SARIMA-GARCH forecasts with those obtained from the best single-method HTS model and the best switching-rule model, respectively; Δ gives the relative advantage of the best switching-rule model compared to the SARIMA-GARCH. Accuracy is presented for selected horizons; the results are preserved in quality for the other horizons. Depending on horizon, the best switching-rule model forecasts 9 - 155 % more accurately than the SARIMA-GARCH model, while the biggest relative advantage of the HTS models is for short-horizon forecasts.

6. Conclusions

We began by using a two-level decomposition of aggregate inflation volatility to shed light on the DGP of inflation volatility. Combining the variance/covariance decomposition with the inflation rate decomposition into common, industry and idiosyncratic components adds to our understanding of the inflation process. The main drivers of the Variance Component (the part of aggregate volatility that is due to variances of product-level inflation rates) are the variances of idiosyncratic shocks. The main drivers of the Covariance Component (the part of aggregate volatility that is due to co-movement of product-level inflation rates) are the variances of the common shocks and the covariances of the product-level idiosyncratic shocks. The industry (product group) level plays only a subsidiary role.

Thus aggregate volatility is generated by a combination of common shocks (driving the covariation of product-level inflation rates) and idiosyncratic shocks (driving the product-level variances). Product-level analysis shows that episodes of high inflation volatility (generally in recessions) are often driven by a single or a few selected products, rather than a sweeping increase in product-level variances – the recession in the early 1990s was characterised by a shock to 'Council Tax & Rates', and covariation of other products with this item; the Great Recession was characterised by a shock to 'Mortgage Interest Payments', and covariation of other products with this item. Both high volatility phases also showed increased covariation of product-level inflation rates in general.

In the second part of the paper, we established the value of hierarchical time series modelling for forecasting the conditional mean and volatility of the aggregate inflation rate. For both the inflation rate and its volatility, the usefulness of explicitly considering the aggregation scheme underlying the RPI is evident from accuracy comparisons against conventional (aggregate) univariate modelling approaches. For the inflation rate (mean model), Middle Out Level 2 ARIMA produces the most accurate forecasts. For inflation volatility (variance model), the results respond to the call for an application of the recently introduced Dynamic Optimised Theta Model to a data set that is dominated by stationarity (by Fiorucci et al., 2016b). Bottom-Up DOTM produces the most accurate forecasts for inflation volatility. This can be attributed to the method's ability to capture local trends very well, and also to the heterogeneity of the Covariance Component's subparts – disaggregated level information that is ignored by aggregate models.

Finally, we presented an extreme variant of the forecast combination approach that appears to work well for inflation forecasting – this involves a dynamic switching-rule that, in a time-varying fashion, applies a combination weight of 1 to the best forecasting model based on an in-sample criterion (in our case the rolling sample variance), and a combination weight of 0 to all other models. While the predicted inflation rate enters inflation volatility through the forecasting system in the standard way (the fitted mean model values are used to obtain the smoothed variance), the model switching approach lets inflation volatility affect the forecast of the inflation rate through the switching rule. This accommodates bi-directional relationship between inflation rate and its volatility. We find that the switching-rule approach can improve forecast accuracy compared to the best single-method HTS model – due to ARIMA models' superior performance in high-volatility episodes, and exponential smoothing-type models' superior performance in low-volatility episodes.

Overall, we find overwhelming support for the use of a hierarchical forecasting approach for UK inflation forecasting. The models can be extended and improved in several ways in future research. First, recent progress in Optimal Combination methods are designed to reconcile individual forecasts at the different levels of disaggregation to make them aggregate consistent – it is to be examined whether using more recently developed approaches can further improve the disaggregated HTS forecasts and/or model switching based forecasts. Second, the UK has experienced major fluctuations in key macro variables over the last three decades. It should be useful to incorporate structural change in forecasting models – for example, by combining the HTS forecasting system with time-varying parameters inflation models – e.g., TVP-VAR models, TVP-FAVAR models, or regime-switching VAR models.[24] Third, the potential of incorporating explanatory variables in the hierarchical forecasting models should be explored – Barnett et al. (2012) find that models

[24] Barnett et al. (2012) argue for the superiority of models with time-varying parameters primarily for the dataset that spans the inflation targeting, a clear structural break. See references therein for an overview of time-varying parameter methods.

that include a large set of explanatory variables tend to do well for quarterly inflation forecasting. Aggregate explanatory variables can easily be incorporated in the HTS framework at least for a subset of forecasting methods. Finally, the issue of choosing a proxy for 'actual' inflation volatility remains open: a possible approach might be to use intra-month dispersion of product prices estimated using daily online prices as in the Billion Prices Project (Cavallo and Rigobon, 2016).

Acknowledgements

The authors thank Andrew Harvey, Vasco Carvalho, and participants of the 36th International Forecasting Symposium, the RSS 2016 International Conference, and the Vienna Congress on Mathematical Finance 2016 for valuable comments and suggestions.

Appendix A – Table of AIC values for candidate GARCH-type models

Table A1 shows AIC values of candidate GARCH-type models. In-mean variants were also estimated, but are not presented due to the insignificance of the in-mean effects in the output equations.

Table A1: AIC Selection of GARCH-type Models.

Model	**AIC**
Assuming Constant Variance of Residuals	
$SARIMA(1,0,1)(1,0,1)_{12}$	-9.082
Adjusting for Non-Constant Variance of Residuals	
$SARIMA(1,0,1)(1,0,1)_{12}$-**GARCH(1,1)**	
Gaussian Error Distribution	-9.388
Student's t Error Distribution	-9.384
Generalized Error Distribution	-9.385
$SARIMA(1,0,1)(1,0,1)_{12}$-**EGARCH(1,1)**	
Gaussian Error Distribution	-9.378
Student's t Error Distribution	-9.374
Generalized Error Distribution	-9.375
$SARIMA(1,0,1)(1,0,1)_{12}$-**CGARCH(1,1)**	
Gaussian Error Distribution	-9.380
Student's t Error Distribution	-9.375
Generalized Error Distribution	-9.376

References

Abadir, K., Talmain, G., 2002. Aggregation, persistence and volatility in a macro model. The Review of Economic Studies 69 (4), 749–779.

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics 135 (1-2), 31–53.

Armstrong, J. S., 2001. Combining forecasts. In: Armstrong, J. S. (Ed.), Principles of Forecasting. Springer, Boston, MA, pp. 417–439.

Armstrong, J. S., 2006. Findings from evidence-based forecasting: Methods for reducing forecast error. International Journal of Forecasting 22 (3), 583–598.
Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: A decomposition approach to forecasting. International Journal of Forecasting 16 (4), 521–530.
Athanasopoulos, G., Ahmed, R. A., Hyndman, R. J., 2009. Hierarchical forecasts for Australian domestic tourism. International Journal of Forecasting 25 (1), 146–166.
Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C., 2011. The tourism forecasting competition. International Journal of Forecasting 27 (3), 822–844.
Atkeson, A., Ohanian, L. E., 2001. Are phillips curves useful for forecasting inflation? Federal Reserve Bank of Minneapolis Quarterly Review 25 (1), 2–11.
Bagdatoglou, G., Kontonikas, A., Wohar, M. E., 2016. Forecasting US inflation using dynamic general-to-specific model selection. Bulletin of Economic Research 68 (2), 151–167.
Ball, L., Cecchetti, S. G., 1990. Inflation and uncertainty at short and long horizons. Brookings Papers on Economic Activity 1/1990, 215–254.
Bandi, F. M., Russell, J. R., 2006. Separating microstructure noise from volatility. Journal of Financial Economics 79 (3), 655–692.
Barnett, A., Mumtaz, H., Theodoridis, K., 2012. Forecasting UK GDP growth, inflation and interest rates under structural change: A comparison of models with time-varying parameters. Working Paper No. 450, Bank of England.
Bean, C., 2003. Inflation targeting: The UK experience. Bank of England Quarterly Bulletin, Winter 2003, 479–494.
Beeson, J., 2016. Consumer price inflation: 2016 weights. Technical Report (22 March 2016), Office for National Statistics.
Belmonte, M., Koop, G., 2013. Model switching and model averaging in time-varying parameter regression models. Working Paper No. 1302, University of Strathclyde Business School, Department of Economics.
Bergmeir, C., Hyndman, R. J., Koo, B., 2015. A note on the validity of cross-validation for evaluating time series prediction. Working Paper No. 10/15, Monash University, Department of Econometrics and Business Statistics.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.
Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69 (3), 542–47.
Box, G. E. P., Jenkins, G. M., 1970. Time Series Analysis: Forecasting and Control. Holden-Day Inc., San Francisco, Calif.
Buncic, D., Moretto, C., 2015. Forecasting copper prices with dynamic averaging and selection models. North American Journal of Economics and Finance 33, 1–38.
Capistrán, C., Constandse, C., Ramos-Francia, M., 2010. Multi-horizon inflation forecasts using disaggregated data. Economic Modelling 27 (3), 666–677.
Carvalho, V., Gabaix, X., 2013. The great diversification and its undoing. American Economic Review 103 (5), 1697–1727.
Cavallo, A., Rigobon, R., 2016. The billion prices project: Using online prices for measurement and research. Journal of Economic Perspectives 30 (2), 151–178.
Chen, H., Boylan, J. E., 2007. Use of individual and group seasonal indices in subaggregate demand forecasting. Journal of the Operational Research Society 58 (12), 1660–1671.
Chen, H., Boylan, J. E., 2009. The effect of correlation between demands on hierarchical forecasting. Advances in Business and Management Forecasting 6, 173–188.
Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting 5 (4), 559–583.
Collins, D. W., 1976. Predicting earnings with sub-entity data: Some further evidence. Journal of Accounting Research 14 (1), 163–177.
Comin, D., Mulani, S., 2004. Diverging trends in macro and micro volatility: Facts. Working Paper No. 10922, National Bureau of Economic Research.
Comin, D., Mulani, S., 2006. Diverging trends in aggregate and firm volatility. Review of Economics and Statistics 88 (2), 374–383.
Cukierman, A., Meltzer, A. H., 1986. A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. Econometrica 54 (5), 1099–1128.
Cukierman, A., Wachtel, P., 1979. Differential inflationary expectations and the variability of the rate of inflation: Theory and evidence. The American Economic Review 69 (4), 595–609.
Dangerfield, B. J., Morris, J. S., 1992. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. International Journal of Forecasting 8 (2), 233–241.
Davis, S. J., Haltiwanger, J., Jarmin, R., Miranda, J., 2006. Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. NBER Macroeconomics Annual 21, 107–156.
Di Fonzo, T., Marini, M., 2011. Simultaneous and two-step reconciliation of systems of time series: Methodological and practical issues. Journal of the Royal Statistical Society. Series C: Applied Statistics 60 (2), 143–164.
Dunn, D. M., Williams, W. H., Dechaine, T. L., 1976. Aggregate versus subaggregate models in local area forecasting. Journal of the American Statistical Association 71 (353), 68–71.
Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.
Faust, J., Wright, J. H., 2013. Forecasting inflation. Handbook of Economic Forecasting 2, 1–56.
Fildes, R. A., Ord, J. K., 2002. Forecasting competitions – their role in improving forecasting practice and research. In: Clements, M. P., Hendry, D. F. (Eds.), Companion to Economic Forecasting. Blackwell, Oxford, pp. 322–353.
Fiorucci, J. A., Louzada, F., Yiqi, B., 2016a. forecTheta: Forecasting Time Series by Theta Models. R package version 2.2.

URL http://CRAN.R-project.org/package=forecTheta
Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., Koehler, A. B., 2016b. Models for optimising the Theta method and their relationship to state space models. International Journal of Forecasting 32 (4), 1151–1161.
Fischer, S., 1981. Towards an understanding of the costs of inflation: II. Carnegie-Rochester Conference Series on Public Policy 15, 5 – 41.
Fliedner, G., 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. Computers and Operations Research 26 (10-11), 1133–1149.
Friedman, M., 1977. Nobel lecture: Inflation and unemployment. Journal of Political Economy 85 (3), 451–472.
Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79 (3), 733–772.
Galati, G., Poelhekke, S., Zhoua, C., 2011. Did the crisis affect inflation expectations? International Journal of Central Banking 7 (1), 167–207.
Gardner, E. S., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4 (1), 1–28.
Gardner, E. S., McKenzie, E., 1985. Forecasting trends in time series. Management Science 31 (10), 1237–1246.
Granger, C. W. J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3 (2), 197–204.
Gross, C. W., Sohl, J. E., 1990. Disaggregation methods to expedite product line forecasting. Journal of Forecasting 9 (3), 233–254.
Hendry, D. F., Hubrich, K., 2006. Forecasting economic aggregates by disaggregates. ECB Working Paper Series No. 589 (Feb 2006), European Central Bank.
Holland, A. S., 1995. Inflation and uncertainty: Tests for temporal ordering. Journal of Money, Credit and Banking 27 (3), 827–837.
Hsiao, C., Wan, S. K., 2014. Is there an optimal forecast combination? Journal of Econometrics 178 (2), 294–309.
Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., Shang, H. L., 2011. Optimal combination forecasts for hierarchical time series. Computational Statistics and Data Analysis 55 (9), 2579–2589.
Hyndman, R. J., Billah, B., 2003. Unmasking the Theta method. International Journal of Forecasting 19 (2), 287–290.
Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 (3), 1–22.
Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22 (4), 679–688.
Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing: The State Space Approach. Springer, Berlin, Heidelberg.
Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.
Hyndman, R. J., Lee, A. J., Wang, E., 2016a. Fast computation of reconciled forecasts for hierarchical and grouped time series. Computational Statistics and Data Analysis 97, 16–32.
Hyndman, R. J., Wang, E., Lee, A., Wickramasuriya, S., 2016b. hts: Hierarchical and Grouped Time Series. R package version 5.0.
URL http://CRAN.R-project.org/package=hts
Jose, V. R. R., Winkler, R. L., 2008. Simple robust averages of forecasts: Some empirical results. International Journal of Forecasting 24 (1), 163–169.
Kahn, K. B., 1998. Revisiting top-down versus bottom-up forecasting. The Journal of Business Forecasting 17 (2), 14–19.
Kim, D.-H., Lin, S.-C., 2012. Inflation and inflation volatility revisited. International Finance 15 (3), 327–345.
Kohn, R., 1982. When is an aggregate of a time series efficiently forecast by its past? Journal of Econometrics 18 (3), 337–349.
Lucas, A., Zhang, X., 2016. Score-driven exponentially weighted moving averages and value-at-risk forecasting. International Journal of Forecasting 32 (2), 293–302.
Makridakis, S., Hibon, M., 2000. The M3-competition: Results, conclusions and implications. International Journal of Forecasting 16 (4), 451–476.
McMillan, D. G., 2014. Forecasting stock returns: Does switching between models help? In: Ma, J., Wohar, M. (Eds.), Recent Advances in Estimating Nonlinear Models: With Applications in Economics and Finance. Springer, New York, pp. 229–248.
Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.
Nikolopoulos, K., Thomakos, D., Petropoulos, F., Litsa, A., Assimakopoulos, V., 2012. Forecasting S&P 500 with the Theta model. International Journal of Financial Economics and Econometrics 4, 73–78.
Office for National Statistics, 2013. Assessment of compliance with the code of practice for official statistics. UK Statistics Authority Assessment Report 246 (March 2013).
Orcutt, G. H., Watts, H. W., Edwards, J. B., 1968. Data aggregation and information loss. The American Economic Review 58 (4), 773–787.
Pegels, C. C., 1969. Exponential forecasting: Some new variations. Management Science 15 (5), 311–315.
Quah, D., 1994. One business cycle and one trend from (many,) many disaggregates. European Economic Review 38 (3-4), 605–614.
Ridge, M., Smith, S., 1991. Local taxation: The options and the arguments. IFS Report Series No. 38, Institute for Fiscal Studies.
Shlifer, E., Wolff, R. W., 1979. Aggregation and proration in forecasting. Management Science 25 (6), 594–603.
Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6), 405–430.
Stock, J. H., Watson, M. W., 2008. Phillips curve inflation forecasts. NBER Working Paper No. 14322, National Bureau of Economic Research.

Taylor, J. W., 2003. Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting 19 (4), 715–725.

Thomakos, D., Nikolopoulos, K., 2014. Fathoming the Theta method for a unit root process. IMA Journal of Management Mathematics 25 (1), 105–124.

Tiao, G. C., Guttman, I., 1980. Forecasting contemporal aggregates of multiple time series. Journal of Econometrics 12 (2), 219–230.

van Erven, T., Cugliari, J., 2015. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In: Antoniadis, A., Poggi, J.-M., Brossat, X. (Eds.), Modeling and Stochastic Learning for Forecasting in High Dimensions. Lecture Notes in Statistics Vol. 217. Springer, Paris, pp. 297–317.

Wickramasuriya, S. L., Athanasopoulos, G., Hyndman, R. J., 2015. Forecasting hierarchical and grouped time series through trace minimization. Working Paper 15/15, Department of Econometrics and Business Statistics, Monash Business School.

Zellner, A., Tobias, J., 2000. A note on aggregation, disaggregation and forecasting performance. Journal of Forecasting 19 (5), 457–465.