Big data forecasting of South African inflation

Byron Botha, Rulof Burger, Kevin Kotze, Neil Rankin, and Daan Steenkamp

ERSA working paper 873

February 2022

Big data forecasting of South African inflation

Byron Botha*, Rulof Burger†, Kevin Kotzé‡, Neil Rankin§, and Daan Steenkamp¶

February 2022

Abstract

We investigate whether the use of machine learning techniques and big data can enhance the accuracy of inflation forecasts and our understanding of the drivers of South African inflation. We make use of a large dataset for the disaggregated prices of consumption goods and services to compare the forecasting performance of a suite of different statistical learning models to several traditional time series models. We find that the statistical learning models are able to compete with most benchmarks, but their relative performance is more impressive when the rate of inflation deviates from its steady state, as was the case during the recent COVID-19 lockdown, and where one makes use of a conditional forecasting function that allows for the use of future information relating to the evolution of the inflationary process. We find that the accuracy of the Reserve Bank's near-term inflation forecasts compare favourably to those from the models considered, reflecting the inclusion of off-model information such as electricity tariff adjustments and within-month data. Lastly, we generate Shapley values to identify the most important contributors to future inflationary pressure and provide policymakers with information about the potential sources of future inflationary pressure.

JEL classification: C10, C11, C52, C55, E31.

Keywords: Micro-data, Inflation, High dimensional regression, Penalised likelihood, Bayesian methods, Statistical learning.

*Codera Analytics. Email: byron@codera.co.za.

†Department of Economics, University of Stellenbosch, Stellenbosch, 7601, South Africa. Predictive Insights, 1st Floor, Rhino House, 23 Quantum Street, Techno Park, Stellenbosch, 7600, South Africa. Email: rulof@sun.ac.za.

‡*Corresponding author:* School of Economics, University of Cape Town, Rondebosch, 7701, South Africa. Predictive Insights, 1st Floor, Rhino House, 23 Quantum Street, Techno Park, Stellenbosch, 7600, South Africa. Email: kevin.kotze@uct.ac.za.

§Predictive Insights, 1st Floor, Rhino House, 23 Quantum Street, Techno Park, Stellenbosch, 7600, South Africa. Email: neil@predictiveinsights.co.za.

¶Codera Analytics and Research Fellow, Department of Economics, Stellenbosch University. Email: daan@codera.co.za.

1 Introduction[1]

Accurate near-term inflation forecasts are an important input into the South African Reserve Bank's (SARB) monetary policy framework. They contribute to an accurate assessment of the economic outlook and therefore an appropriate policy stance. For this reason, it is important that the SARB draws on all available information and compares the performance of different forecasting approaches. Recent advances in the field of statistical learning have facilitated a number of investigations that utilise different types of big data.[2] Such investigations are also of importance to policymaking institutions, where several central banks are making use of these techniques to conduct economic research and inform policy decisions.[3] In addition, studies that make use of these methodologies have also influenced those policy decisions that consider the impact of the COVID-19 pandemic on economic activity.[4] In this paper, we make use of both aggregated and disaggregated data on consumer prices that is collected by Statistics South Africa (StatsSA) to construct the South African Consumer Price Index (CPI) and a large suite of statistical models to predict future measures of headline and core inflation.[5] Since the expected value of this variable influences many different types of policy decisions, such as the future rate of interest, it would potentially affect the activity of a diverse set of economic agents. In addition, such forecasts anchor inflation expectations, which may improve policy efficacy and economic stability, to provide an improved foundation for higher levels of economic growth.

To aid the discussion on the relative performance of the different models that are used to generate the respective forecasts, we make use of four broad categories. The first of these relate to the *benchmark models*, which include traditional random walk, autoregressive and Bayesian vector autoregressive (BVAR) specifications. In addition, we also include the forecasts from the central bank's disaggregated inflation model (DIM), which is largely responsible for influencing the monthly near-term inflation forecasts, along with the monthly inflation forecasts that are presented to the monthly Monetary Policy Committee (MPC) meetings. The second group of models make use of *dimensionality reduction* techniques that seek to summarise all of the data from the potential predictors and would include those frameworks based upon principal component analysis. The third group of models make use of *variable selection* techniques and would include models that make use of shrinkage estimators, penalised likelihood functions or Bayesian model selection techniques. And then finally, the fourth group of models include the use of *non-linear statistical learning* forecasting models, such as the Random Forrest and Neural Network, which may also incorporate non-parametric features.

This paper is related to a number of different strands in the literature. The first of these includes the comparative

[1] The authors would like to thank Patrick Kelly and Marietjie Bennett from Statistics South Africa for their assistance with the data. We are also grateful for the comments from David Fowkes and an anonymous referee.

[2] Agrawal et al. (2019), Athey (2017, 2018), Athey and Imbens (2019), Mullainathan and Spiess (2017) and Varian (2014) contain overviews of selected research and discussions relating to the potential use of statistical learning methods within the field of economics.

[3] In a recent survey, Doerr et al. (2021) note that $\pm$ 80% of central banks discuss the topic of big data formally, where 70% of these monetary authorities use it for economic research, while 40% use it to inform policy decisions. Around two thirds of respondents indicated that they wanted to start new big data projects in 2020/21. The results from the study also suggest that the number of central bank speeches that mention the use of big data has increased significantly over recent periods of time and that most do so in a positive light. For an earlier discussion on the use of statistical learning methods within central banks and other policymaking institutions, see Wibisono et al. (2019), Tissot (2019), Mehrhoff (2017), Hammer et al. (2017), Baldacci et al. (2016), Florescu et al. (2014), World Bank (2014) and United Nations Global Pulse (UNGP) (2012).

[4] For example, Blumenstock (2020) describes a few practical cases where the use of these techniques may be applied in developing countries, while the Organisation for Economic Co-operation and Development include mention of ways in which these techniques may be used to identify potential responses that may ease the effects of the pandemic (OECD, 2020). In addition, Buckman et al. (2020) make use of the method that is employed in Shapiro et al. (2017) to report on changes in consumer sentiment following the onset of the pandemic, while Chetty et al. (2020) construct daily indices on consumer spending and other indicators disaggregated by zip code, industry and income to show that high-income households reduced spending by more than low-income households, which has contributed towards job losses among the low-income households that provide services to high-income households. Similar changes in consumption behaviour have been noted in Baker et al. (2020b), who make use of transaction-level financial data to explore how household consumption responded to the onset of the pandemic, while Baker et al. (2020a) report on the changes in uncertainty relating to consumption spending. Other research by Chakrabarti et al. (2020a,b) make use of large datasets to investigate changes in consumer spending and business revenue in response to state re-openings, while Carvalho et al. (2020) note that in Spain, the consumption baskets converge towards the goods basket of low-income households. Similar changes in the consumption basket over this period of time have also been observed in Cavallo (2020) for a number of countries.

[5] Previous studies that have used the disaggregated consumer price survey data for assessing pricing behaviour in South Africa include Creamer and Rankin (2008), Creamer et al. (2012), Ruch et al. (2016a), and Ruch et al. (2016b). This dataset includes prices for 34,075 unique goods and services, which through various methods of aggregation is reduced to a set of 216 predictors for the period January 2009 to March 2021. Hence, it incorporates 12 months over which various lockdown measures were imposed.

studies that consider different models that may be used to forecast inflation. For example, Faust and Wright (2013) suggest that judgement forecasts, such as those from the Federal Reserve or inflation expectation surveys, tend to be more accurate than forecasting model predictions, while Stock and Watson (2007) suggest that relatively simple univariate models provide good comparative forecasts of US inflation, where a model that incorporates both unobserved components and stochastic volatility provide a reasonably good forecast of inflation.[6] These findings, which largely relate to a period of stable economic activity for a developed economy, should not be too surprising as the conditional mean of inflation was highly persistent (Fuhrer, 2010; Wolters and Tillmann, 2015). However, over periods where economic activity is not particularly stable, a forecast that is close to the previously observed mean may not be terribly accurate and the models may need to allow for sustained departures from steady-state values. Hence, it may be necessary to make a number of amendments to traditional forecasting models during a period of economic crisis, or where the rate of inflation is relatively variable, as in the case of several low- and middle-income countries.

The paper also seeks to contribute towards the literature that considers the relative performance of inflation forecasting models in an emerging country such as South Africa, which include those that emphasise the structural features of an economy. For example, in an early study, Woglom (2005) notes that inflation forecasts that are generated from a simple Phillips curve are not particularly accurate. However, when making use of a more expansive variant of a structural model, Smal et al. (2007) suggest that such models are capable of producing quarterly forecasts for CPIX[7] inflation that are more accurate than either the autoregressive integrated moving average or disaggregate inflation model. In addition, these forecasts were also shown to be more accurate than the Reuters consensus forecast over their particular sample. Subsequent structural models, which include Liu et al. (2009), suggest that the forecasts from a small closed-economy New Keynesian dynamic stochastic general equilibrium (NKDSGE) model outperform those that are generated by classical vector autoregressive (VAR) and BVAR models for the South African GDP deflator. However, these authors also note that the differences in root-mean squared-error (RMSE) was not significant. Thereafter, Steinbach et al. (2009) extended the NKDSGE model to incorporate small open-economy features and found that the model's forecasts for CPIX inflation provided a lower RMSE, when compared to the Reuters consensus forecast, over a horizon that extends between four and seven quarters ahead. Similarly, Alpanda et al. (2011), built upon the small open-economy NKDSGE model that is discussed in Alpanda et al. (2010a) and Alpanda et al. (2010b), to show that their model provides better forecasts for consumer price inflation over shorter horizons. Furthermore, they also show that the difference in performance relative to classical VAR, BVAR and random walk models is significantly different from zero.

This literature has subsequently been extended to consider the performance of a small open-economy NKDSGE-VAR model in Gupta and Steinbach (2013), which generates CPIX inflation forecasts that are superior to classical VAR and most BVAR models (with the exception of a BVAR model that incorporates a stochastic search variable selection prior) over a one-quarter ahead horizon. While other researchers have considered the role of nonlinearities within structural models, where Balcilar et al. (2015) make use of a nonlinear NKDSGE model, which employs the second-order solution method of Schmitt-Grohé and Uribe (2004) and a particle filter to evaluate the likelihood function, to provide forecasts for consumer inflation that have a lower RMSE, when compared to a large variety of BVAR models (including those that employ variable selection priors). Furthermore they found that the difference in forecasting performance is usually statistically significant, when compared to a random walk and linear NKDSGE model (particularly over longer horizons). However, when considering the use of regime-switching nonlinearities, Balcilar et al. (2017) note that the out-of-sample forecasts for South African inflation that are generated by different forms of Markov-switching NKDSGE models are largely inferior to the single regime counterpart.

There are also a number of papers that focus on the application of different non-structural statistical techniques

[6] In addition, for the period that incorporates the financial crisis, Stock and Watson (2010) suggest that this model should incorporate a stochastic trend that reacts to the unemployment recession gap, where the short-term response of inflation is consistent with an increase in this gap, while the long-term response is dependent upon the persistence in trend inflation. As is the case with most low- and middle-income countries, South Africa does not have a reliable measure for the unemployment recession gap that could be applied in an investigation that makes use of monthly observations of time series variables.

[7] A measure of consumer price inflation that excludes the effects of interest rates on mortgage bonds.

to forecast South African inflation, to which this paper contributes. For example, in an attempt to reduce the potential effects of an omitted variable bias, Gupta and Kabundi (2011) make use of the Stock and Watson (2002b) and Forni et al. (2000) large factor models to forecast the percentage change in the implicit GDP deflator, along with the percentage change in real per capita GDP and the 91-day Treasury Bill rate in South Africa, over a one- to four-quarter ahead period from 2001Q1 to 2006Q4. They make use of 267 quarterly macroeconomic series to show that the factor models tend to outperform the unrestricted VAR, BVAR and small closed-economy NKDSGE models. Similar results are provided in Gupta and Kabundi (2010), where it is noted that large-scale data-rich models are better suited to forecasting key macroeconomic variables, relative to small-scale models. As an alternative, Kanda et al. (2016) is one of the few studies that make use of monthly data to focus on evaluating the performance of a suite of univariate non-linear models, which include a locally linear model tree, neuro-fuzzy, multi-layered perceptron, artificial neural network, non-linear autoregressive, and genetic algorithm based forecasting model. Their findings suggest that the locally linear model tree provides forecasts that can compete with the linear autoregressive model and is generally superior over longer horizons. In addition, Ruch et al. (2020) derive forecasts for quarterly measures of core inflation in South Africa with the aid of time-varying parameter vector autoregressive models (TVP-VARs), factor-augmented VARs, and structural break models to show that small TVP-VARs outperform all their other models, where additional information on the growth rate of the economy and the interest rate is sufficient to forecast core inflation accurately.

More recently, Galvao (2021) summarises a number of recent developments from the international literature, which seek to address some of the challenges that may arise when looking to forecast macroeconomic variables during a pandemic, where Castle et al. (2021) and Goulet Coulombe et al. (2021) note that statistical models or nonlinear machine learning models that are able to adapt to various changes, may perform better than well-specified structural models, while Koop et al. (2021) suggest that modelling specifications that accommodate time variation in forecasting uncertainty may also provide improved results.

While the SARB's model of near-term inflation includes the 46 monthly components used to measure headline CPI, this paper estimates a suite of big data models that draw on data at a higher level of disaggregation (at four-, five- and eight-digit classifications). A four-digit classification, for example, considers bread and cereal categories instead of brown bread, which is an eight-digit classification. Our results suggest that the combined use of big data and statistical learning methods provide results that are potentially able to compete with certain benchmarks over longer horizons, however, many of the traditional benchmarks are superior over shorter horizons. Specifically, we find that the accuracy of SARB's near-term inflation forecasts compare favourably to those from the models considered, reflecting the inclusion of off-model information, such as electricity tariff adjustments and within-month data. Furthermore, we also note that the relative performance of the statistical learning methods is more impressive when the rate of inflation deviates from its steady state during the period of the economic lockdown and where we make use of a conditional forecasting function that allows for the use of future information relating to the evolution of the inflationary process.

The paper is also related to a second strand of literature, which considers the idiosyncratic behaviour of consumer prices, where the frequency and dispersion of price adjustments can vary across items and over time (Chu et al., 2018; Petrella et al., 2019; Stock and Watson, 2020; Chetty et al., 2020; Carvalho et al., 2020; Cavallo, 2020). Given these characteristics of the data we could conceive that when the price indices are subjected to various forms of aggregation, their predictive power may decline. For example, if the disaggregated price index for brown bread has impressive predictive power, while the other products in category for breads and cereals are poor predictors then the signal that is provided by brown bread may be obscured if we were to restrict the analysis to use the aggregate data for the category rather than the individual goods. Previous findings in Hubrich and Hendry (2005) suggest that the use of disaggregated CPI components for the United States does not result in a meaningful improvement in forecasting accuracy, while studies that were conducted for Mexico and Portugal suggest that the use of disaggregated components could provide notable improvements (Ibarra, 2012; Duarte and Rua, 2007). After making use of data at different levels of aggregation we show that when using higher levels of disaggregated data, the forecasting results for headline inflation do indeed improve over all horizons.

The other strand of literature to which this paper contributes, involves the relative merits of employing techniques that seek to summarise all the available information (which is contained in the set of potential predictors), as opposed to only selecting those variables from a set of potential predictors that provide useful predictive power (i.e. the *density* versus *sparsity* debate). Giannone et al. (2021) have suggested that when making use of various macroeconomic and financial data sets for the United States, the forecasts of dense models are more accurate than the sparse counterparts, while Joseph et al. (2021) note that when restricting the set of potential regressors to disaggregated consumption price indices for the United Kingdom, the sparse models provide more impressive results.[8] Our results suggest that forecasts of several sparse models are superior to those of the dense models, when making use of lower levels of price aggregation. For example, both the least absolute shrinkage and selection operator (LASSO) and the ridge regression provide results that are superior to the dynamic factor models over most horizons when using data for headline inflation at the 8/5 digit level of disaggregation. Hence, there would appear to be advantages to identifying those variables that contribute towards the underlying predictive signal in the data, to restrict the information that is used in the construction of the forecast to those variables that have substantive predictive power.

There is also a fourth strand of the literature that is related to this paper, which considers the use of non-linear statistical learning models that are able to learn unknown functional forms and potential structural changes in both the mean and trend. Medeiros et al. (2021) suggest that the use of these methods is superior to those of alternative methodologies over the medium to longer horizons, when making use of the large macroeconomic dataset for the United States (the construction of which is described in McCracken and Ng (2016)). Our results suggest that this finding would only hold when we make use of the conditional forecasting function that is used in Medeiros et al. (2021). For example, when making use of future information about the value of inflation, the neural network, random forests and gradient boosting models all provide some of the most impressive results for both core and headline inflation over longer horizons, however, when using an unconditional forecasting function these models are no longer able to generate superior forecasts.

In the penultimate part of the analysis, we derive Shapley values from the forecasts to identify the relative contribution of each of the disaggregated predictors to the forecast. These results contribute towards the literature that considers the use of disaggregated consumer data and the role of price setting behaviour in the identification of the current state of the business cycle, which is summarised in Klenow and Malin (2010). Such studies may be used to inform macroeconomic models, which make use of rigidities in the pricing mechanism to allow for monetary policy to have real effects. More recently, Nakamura and Steinsson (2013) considered a number of important features of the price adjustment mechanism, which may influence the economy's response to demand shocks, while in the case of South Africa, Ruch et al. (2016a) and Ruch et al. (2016b) have sought to identify the goods and services that are responsible for generating inflation. This investigation is also related to the work that has been conducted on the identification of various elements that should be incorporated in core inflation, which could be defined as the measure of inflation that excludes those items that are subject to large transitory movements that may not have a significant effect on the long-term future path of prices, as in Du Plessis et al. (2015).

The remaining sections of this paper are organised as follows. Section 2 describes the methodology of the various models that have been specified in this study, while details relating to the data are discussed in section 3. The results of the various investigations that have been conducted are presented in section 4, before section 5 concludes.

2 Methodology

To describe the methodology that has been employed by the various models it is necessary to introduce some notation. In all that follows, we assume that $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ is a vector of data for the measure of inflation, where

[8] Joseph et al. (2021) also find that after incorporating additional measures of macroeconomic activity, the dense models then provide more accurate forecasts, which support the findings of Giannone et al. (2021).

the observations that arise over time are denoted, $i \in \{1,\ldots,n\}$. The matrix for the set of predictors that include the price indices for the different products or categories that are sampled to construct the CPI are contained in $\boldsymbol{X} = \{x_{1,1},\ldots,x_{n,p}\}'$, which is of dimension $(n \times p)$, while $j \in \{1,\ldots,p\}$ is used to denote each of the different predictors in the matrix. To consider the relative forecasting accuracy of the different models, we make use of a conditional and unconditional recursive out-of-sample investigation that extends over a horizon of between one and 24 months ahead, where the data that is used to test the predictions extends over a four-year period. The motivation for making use of a recursive forecasting scheme, as opposed to a rolling-window scheme, is that we do not have a large number of available observations that have been measured over time. For example, if we were to make use of a rolling-window scheme, then we would have been limited to making use of a constant in-sample period for the predictors of just over five years to generate a 24-month ahead forecast, when using most of the statistical learning models. Since we have a large number of potential predictors, we have assumed that by making use of a slightly larger in-sample dataset, we could possibly generate more accurate forecasts for the observations that arise over more recent periods of time. The statistics that are used to evaluate the out-of-sample performance of the respective models include the root-mean squared-error (RMSE), the mean absolute percentage error (MAPE), and the Diebold and Mariano (1995) statistics.[9] When reporting on the results, we consider the year-on-year forecasts of headline and core inflation.

2.1 Benchmarks and dynamic factor models

To evaluate the relative forecasting performance of the statistical learning models, we consider the use of a number of benchmarks, which are provided by autoregressive, large-scale Bayesian vector autoregressive, stochastic volatility, and random-walk models. Additional benchmarks include the model that is currently used by the central bank in South Africa to generate short-term monthly inflation forecasts, and the actual forecasts that are presented to the MPC. The latter incorporate off-model information such as electricity tariff adjustments and within-month data outturns. Where there are relatively few predictors, we also include the results from a linear regression model. To compare the results of models against a suite of dense models that make use of principal components to summarise linear combinations of the original predictors. Two different variants of the dynamic factor model (DFM) are employed, where the first builds on the traditional DFM, which largely follows the seminal work of Forni et al. (2000), Stock and Watson (2002a,b) and Bai (2003), and makes use of the target factor approach that follows the work of Bai and Ng (2008), while the second approach utilises the three-pass regression filter of Kelly and Pruitt (2013, 2015).[10]

2.2 Statistical learning variable selection models

Within the context of a linear regression model, when $p \to n$ the parameters are estimated with decreasing degrees of precision and when $p > n$ then the parameter estimates are not unique. In such cases, the use of the maximum likelihood estimate is no longer appropriate, and as an alternative, we could make use of a penalised likelihood function, where we impose a cost for including individual covariates within the $\boldsymbol{X}$ matrix.

[9] Although there will be cases where these models will be nested, which would imply that the use of the statistics that are discussed in Clark and West (2007) and McCracken (2007) would be preferred to the Diebold and Mariano (1995) statistic, these models are not always nested within one another. Therefore, for the sake of consistency, we make use of the Diebold and Mariano (1995) statistic to evaluate all of the models. This would suggest that the results may favour the more parsimonious model in those cases where the models are nested.

[10] Further details relating to the specification of the benchmark models is included in section A of the appendix, while section B of the appendix contains details that pertain to the models that employ techniques that summarise the information that is provided by the predictors to reduce the dimensionality of $\boldsymbol{X}$.

Such a likelihood function could take the form,

$$\arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^{\top} \boldsymbol{\beta} \right)^2 + h\left(\boldsymbol{\beta}\right) \tag{1}$$

where $h(\boldsymbol{\beta})$ is a penalty function that is applicable when $\hat{\beta}_j \neq 0$, which is usually specified with the aid of L_1 penalties (that are imposed on the absolute value of the number of entries in $\hat{\boldsymbol{\beta}}$), L_0 penalties (that are imposed on the number of non-zero elements that arise in $\hat{\boldsymbol{\beta}}$), L_2 penalties (that seek to shrink those elements in $\hat{\boldsymbol{\beta}}$ that are not estimated with sufficient precision towards zero), or folded concave penalties (that make use of various combinations of the above methods).

If we were to assume that there are several elements within $\hat{\boldsymbol{\beta}}$ that take on values that are close to zero, then we may want to remove the associated predictors from $\boldsymbol{X}$, as this would reduce the estimation error, which may result in an improved bias-variance trade-off. In what follows, we make use of the notation $\boldsymbol{\beta}^*$ for the true value of the coefficients, such that $E\left[\boldsymbol{y}\right] = \left(\boldsymbol{X}\boldsymbol{\beta}^*\right)$, where p^* denotes the number of non-zero values in $\boldsymbol{\beta}^*$.

2.2.1 Least absolute shrinkage and selection operator

Using the method that was initially proposed by Tibshirani (1996), we may look to impose an L_1 penalty that takes the following form:

$$\hat{\boldsymbol{\beta}}^{(\lambda)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^{\top} \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \tag{2}$$

where λ is a hyper-parameter that influences the size of the penalty, such that when $\lambda = 0$ then we have the traditional least-squares estimator, while when $\lambda \to \infty$ then most of the $\hat{\beta}_j^{(\lambda)}$ elements would take on a value of zero. In our case, we apply cross-validation methods to identify an appropriate value for λ, where we make use of ten folds of the data.[11]

This estimator is termed the LASSO and Hastie et al. (2015) note that the computational time that is taken to derive the likelihood function for this estimator in both linear and generalised linear models is relatively small. Furthermore, they show that it provides reasonable predictions for $\boldsymbol{y}$, when we have up to $\left(p^* \log(p) \ll n\right)$ co-variates. However, it has also been noted that the estimates for $\hat{\boldsymbol{\beta}}$ would contain a non-negligible bias and in a number of empirical applications it has been suggested that this method has a tendency to select too many predictors.

2.2.2 Adaptive LASSO

There are several variations of L_1 methods, which include the use of folded concave penalties, such as the adaptive LASSO model of Zou (2006). To implement this procedure, we obtain preliminary estimates for $\hat{\beta}$ after making use of the traditional LASSO routine. Thereafter, we make use of a similar LASSO routine with the set of variables that were deemed to have reasonable predictive power, but in this case we divide the penalty for the β_j estimates by $\hat{\beta}_j$, which was obtained from the first estimation. Therefore, the second estimation makes use of the following minimisation routine:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^{\top} \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} \frac{\left| \beta_j \right|}{\left| \hat{\beta}_j \right|} \tag{3}$$

[11] As an alternative we could have made use of information criteria for this purpose.

This would imply that the number of regressors that will remain after the second estimation will be proportional to the number that remain after the first. So if the value for λ is relatively large following the first estimation, then it will be relatively small for the second estimation, and vice-versa.

From a computational perspective, this requires that we run the estimation routine on two occasions, but the second estimation should be relatively expedient since the number of variables that have predictive power after the first estimation is usually quite small. Of course this routine would reduce the number of regressors that are deemed to have predictive power, which is desirable as we have already noted that the traditional LASSO tends to over-select, and by applying the methodology of the adaptive LASSO we reduce the bias, as $\hat{\beta}_j$ increases in size.

2.2.3 Post-selection inference

The procedures for uncertainty quantification, when making use of a penalised likelihood function, are not particularly straightforward, and as such, the use of traditional methods for the calculation of confidence intervals and associated probability values for the significance of coefficients may not be appropriate.

One way to proceed would be to run an ordinary-least-squares regression on the variables that have been selected by the model that makes use of a penalised likelihood function. However, such results are based on the assumption that the penalised likelihood function has identified the correct predictors, which might not necessarily be the case. Furthermore, the result of this procedure would suggest that the probability values would be overly liberal and potentially biased, as we are making use of a repeated sampling procedure on a single dataset. To alleviate such concerns one could split the dataset into different samples, but where the number of observations is limited, this may affect the precision of the coefficient estimates. Note also that in a nested model setting, it can be shown that the confidence intervals are smaller when we have a more parsimonious model (i.e. one that makes use of a smaller number for p).

To illustrate one of the solutions that has been proposed for this problem, consider a setting where the covariates are uncorrelated (i.e. $X^\top X = nI$) and none of the explanatory variables have an effect on $\boldsymbol{y}$ (such that $\beta_j^* = 0$). In such a case, the ordinary-least-squares estimates, $\hat{\boldsymbol{\beta}}$, would have a sampling distribution of $\hat{\beta}_j \sim N\left(0, \sigma^2/n\right)$. However, when the parameter estimates are derived from a LASSO procedure, for $\hat{\boldsymbol{\beta}}^{(\lambda)}$, we would only have coefficient estimates for a subsample of the possible regressors, where $|\hat{\beta}_j| > \lambda$. This would imply that the sampling distribution for $\hat{\boldsymbol{\beta}}^{(\lambda)}$ could be described by a truncated normal distribution that takes the form:

$$\hat{\boldsymbol{\beta}}^{(\lambda)} | \beta_j^* = 0 \sim N(0, \sigma^2/n) I\left(|\hat{\beta}_j| > \lambda\right) \tag{4}$$

More generally, in a setting where the $\boldsymbol{X}$ matrix contains correlated variables, Lee et al. (2016) show that the sampling distribution of $\hat{\boldsymbol{\beta}}^{(\lambda)}$ is given by a multivariate normal distribution with polyhedral constraints that makes use of a similar argument. Hence, when making use of ordinary-least-squares regression, the usual formula for a 95% confidence interval would be $\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)$, where the 1.96 is the 0.025 quantile of a $N(0,1)$ distribution. To calculate an equivalent value for the truncated normal distribution, let $Z \sim N(0,1),\ W \sim N(0,1)|(W > \lambda)$. Then for any $w > -\lambda$, the cumulative distribution function would be:

$$P(W < w) = \frac{P(Z < w)}{2P(Z < w)} \tag{5}$$

Now if λ is such that the truncated region has a probability region that is equivalent to $1/4$ of the normal density, then ± 1.96 should be replaced by:

$$\hat{\beta}_j \pm 1.96\frac{4}{3} \times SE\left(\hat{\beta}_j\right) = \hat{\beta}_j \pm 2.61 \times SE\left(\hat{\beta}_j\right) \tag{6}$$

More generally, this adjustment requires solving an optimisation problem that calculates the necessary degree of truncation, but in almost all cases, when performing an ordinary-least-squares regression after obtaining the LASSO estimates, the confidence intervals should be wider as we would need to be more conservative.[12] In our setting, we make use of post-selection inference to exclude those predictors that do not have significant explanatory power. For example, if the LASSO procedure were to suggest that to maximise the likelihood function we should include a particular predictor that is largely unable to explain the future evolution of inflation, then the post-selection inference procedure could be used to remove this particular predictor from the variables that will be used to generate the forecast.

2.2.4 Post-LASSO estimation

Belloni et al. (2011, 2013, 2014, 2017) describe the use of a post-LASSO estimation procedure that has impressive properties in a number of different settings, which include the use of instrumental variables, various control variables, quantile regression methods, etc. In particular, we follow Belloni and Chernozhukov (2013) and make use of the least-squares post-LASSO estimator, which seeks to minimise the least squares criterion over the non-zero components that are selected by the initial LASSO estimator. This procedure involves the initial estimation of a sparse regression model, such as a LASSO, where the authors make use of a specific data-driven penalty that they have specifically derived for this purpose. Thereafter, they apply a duly amended ordinary least squares regression to the model variables that were selected from the first step. The intuition for this result is that the initial LASSO-based model selection would usually omit those components that have relatively small coefficients, while the post-LASSO estimation would remove additional components that may have slightly larger coefficients, but where such a coefficient is insignificantly different from zero.

To ensure that this estimator remains valid, despite the use of a two-stage procedure Belloni and Chernozhukov (2013) show how the model parameters would need to be estimated. In addition, they also provide theoretical results, which suggest that the application of this strategy would provide a smaller bias in situations that include cases where the LASSO-based model does not identify the best-dimensional approximation of the "true" regression function.

2.2.5 L_0 penalties

As an alternative to the above, we may elect to impose the L_0 penalty on the model, which has improved properties over the above L_1 methods. In this case we make use of the term $|\boldsymbol{\beta}|_0$, which is the number of non-zeroes in $\boldsymbol{\beta}$, such that, $|\boldsymbol{\beta}|_0 = \sum_{j=1}^{p} \Psi\left(\beta_j \neq 0\right)$, where Ψ is an indicator function that takes on a value of 1 when $\beta_j \neq 0$. In other cases it takes on a value of 0. Imposing the L_0 penalty on the model would then involve solving the following minimisation problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda|\boldsymbol{\beta}|_0 \tag{7}$$

To select the correct value for λ we could make use of information criteria, where the Akaike Information Criteria (AIC) would set $\lambda = 2|\boldsymbol{\beta}|_0$, while the use of Bayesian Information Criteria (BIC) would set $\lambda = \log(n)|\boldsymbol{\beta}|_0$. When comparing these two methods, several studies have found that the use of the AIC would usually result in the inclusion of too many elements in $|\boldsymbol{\beta}|_0$. In addition, when working with two nested models with $p_1 < p_2$ variables, it has been shown that the AIC does not consistently select the true model, while the BIC does, provided that $(p_2 - p_1)\log(n) \ll n$.

[12] Additional literature on the construct confidence intervals for selected $\hat{\beta}_j \neq 0$ in this setting is contained in Taylor and Tibshirani (2018).

In cases where $p > n^2$, researchers would usually make use of the extended BIC, where $\lambda = \log(n)|\boldsymbol{\beta}|_0 + \log\binom{p}{|\boldsymbol{\beta}|_0} \approx \log\left(np^2\right)|\boldsymbol{\beta}|_0$. Rossell (2021) has suggested that when imposing L_0 penalties, under the condition that $n \to \infty$ and where k^* is the true model, we would need to solve:

$$\lim_{n\to\infty} P\left(\text{EBIC}_{k^*} < \min_{k\neq k^*} \text{EBIC}_k\right) = 1 \tag{8}$$

To a high degree of approximation, these methods may be imposed with the aid of the Bayesian model selection framework and as such all the computational tools, which include Markov Chain Monte Carlo methods, that are prevalent in the Bayesian literature may be applied to this setting. Common methods to search the model space include the use of the stepwise forward routine (where we start with no covariates), stepwise backward (where we start with $\bar{p}$ covariates), some combination that allows for forward and backward moves, or blockwise methods (where we add or drop groups of highly-correlated variables).

Although computation with the aid of L_0 penalties is generally slower, there are small prediction benefits that may be derived when seeking to generate values for $\boldsymbol{y}$, and we are also able to work with a large number of potential predictors (that could be of dimension, $p^* \log(p) \ll n$). For the purposes of identification, when making use of a well-chosen value for λ and where the number of elements in $\hat{\boldsymbol{\beta}}$ is large, these methods have less bias than the LASSO and it is more likely to select the true set of predictors (particularly when there is a large degree of correlation in $\boldsymbol{X}$).

2.2.6 Ridge regression

Parameter shrinkage or L_2 penalties may be incorporated within the specification of a number of different models. One of the more popular are termed ridge regressions, which seek to adjust coefficient estimates in linear models, where those that are insignificantly different from zero, would tend towards values of zero. The objective of this procedure would be to provide more accurate coefficient estimates for those coefficients that are different from zero. This particular model makes use of shrinkage by placing a constraint on the sum of squared β^2 parameter values:

$$\hat{\boldsymbol{\beta}}^{(\lambda)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{9}$$

where λ influences the degree of shrinkage and a larger value of λ would exert more influence over directing the coefficient estimates towards zero. In our case, we make use of cross validation techniques to determine the size of λ. Note that in contrast with L_1 and L_0 penalties, in this case we may have a number of coefficients that are close to zero (but not exactly equal to zero).

2.2.7 Elastic Net

Alternative specifications of penalties may use combinations of both L_1 and L_2 penalties. One example is provided by the elastic net, which seeks to make a compromise between the ridge regression and LASSO penalty. Such a specification could take the form:

$$\hat{\boldsymbol{\beta}}^{(\lambda)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda_2 \sum_{j=1}^{p} \left|\beta_j\right| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{10}$$

Given this specification, additional parameters could be used to influence the prevalence of each penalty. For example, we could apply a weight of α to the L_1 penalty term and a weight of $(1-\alpha)$ to the L_2 penalty term, where α takes on values between 0 and 1. In our case we have implemented two variants of this model to

consider the use of two different L_1 penalties, which include both the traditional LASSO and the adaptive LASSO penalty. In addition, we have also incorporated the results for the smoothly clipped absolute deviation (SCAD) penalty of Fan and Li (2001), which seeks to make use of a similar methodology. However, in this case, large values of $\hat{\boldsymbol{\beta}}$ are left alone, small values are set to zero, while intermediate values are shrunk towards zero.

2.2.8 Bayesian model selection

In addition to the frequentist techniques that have been discussed above, there are also a number of Bayesian methods that may be used to identify the most relevant variables within a high-dimensional regression problem. As per the standard Bayesian approach, these methods combine prior knowledge about the model parameters, $p\left(\boldsymbol{\beta},\sigma^2\right)$, along with the likelihood function, $p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right)$. This provides the posterior distribution, $p\left(\boldsymbol{\beta},\sigma^2|\boldsymbol{y}\right)$, which contains information about the model parameters, $\boldsymbol{\beta}$, after observing the data. Hence, we could summarise this approach to parameter estimation as follows:

$$p\left(\boldsymbol{\beta},\sigma^2|\boldsymbol{y}\right)=\frac{p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right)p\left(\boldsymbol{\beta},\sigma^2\right)}{p\left(\boldsymbol{y}\right)}\propto p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right)p\left(\boldsymbol{\beta},\sigma^2\right) \tag{11}$$

where $p\left(\boldsymbol{y}\right)=\int p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right)p\left(\boldsymbol{\beta},\sigma^2\right)d\boldsymbol{\beta}d\sigma^2$ ensures that the probability distribution integrates to unity. If we focus our attention on the parts of the model that incorporate the parameters, we would note that the posterior distribution is proportional to the product of the prior distribution and the likelihood function. If p is relatively small and as $n\rightarrow\infty$ then $p\left(\boldsymbol{\beta}|\boldsymbol{y}\right)$ converges to the sampling distribution of the maximum likelihood estimate under minimal regularity conditions. However, if $p\gg n$, this is not necessarily the case.

In the linear regression model the likelihood function assumes that $\boldsymbol{y}$ takes a multivariate normal distribution that is related to $\boldsymbol{X}\boldsymbol{\beta}$, where the residual has a variance of σ^2, such that $p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right)=N\left(\boldsymbol{y};\boldsymbol{X}\boldsymbol{\beta},\sigma^2I\right)$. In these settings a convenient prior, $p\left(\boldsymbol{\beta}|\sigma^2\right)$, would take the form of a multivariate normal distribution, $N\left(\boldsymbol{0},\sigma^2\boldsymbol{A}\right)$. Hence, for $\boldsymbol{\beta}\in R^p$, we would have $\boldsymbol{\beta}\sim N\left(\mu,\Sigma\right)$, where the properties of the mean may be summarised as, $\mu\in R^p$, and the covariance, Σ, is of dimension $p\times p$, such that:

$$\left(\boldsymbol{\beta}|\mu,\Sigma\right)=\frac{1}{\left(2\pi\right)^{\frac{p}{2}}\left|\Sigma\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)^{\top}\Sigma^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)\right\} \tag{12}$$

The prior for the residual variance, $p\left(\sigma^2\right)$, is often assigned an inverse gamma distribution, $IG\left(a,b\right)$, which is summarised by the probability density function:

$$p\left(\sigma^2|a,b\right)=\frac{b^a}{\Gamma\left(a\right)}\left(\frac{1}{\sigma^2}\right)^{a+1}e^{-b/\sigma^2} \tag{13}$$

This particular distribution is relatively flexible and generalises the inverse exponential and χ^2 distributions, where $E\left(\sigma^2\right)=b/\left(a-1\right)$ for $a>1$, and $var\left(\sigma^2\right)=\frac{b^2}{(a-1)^2(a-2)}$ for $a>2$. To derive an interpretation for the values of a and b, note that when making use of a relatively uninformative prior, where $\sigma^2\sim IG\left(0.01,0.01\right)$, it would be equivalent to making use of information from a prior study, where $n=0.01$, and $\hat{\sigma}=1$.

The value for $\boldsymbol{A}$ in the multivariate normal distribution, $N\left(\boldsymbol{0},\sigma^2\boldsymbol{A}\right)$, may take the form of a diagonal matrix, or alternatively, we could make use of the prior that was developed in Zellner (1986), where $\boldsymbol{A}=\boldsymbol{X}^{\top}\boldsymbol{X}$, such that a prior experiment with 1 observation provides an estimated $\hat{\boldsymbol{\beta}}$ of zero, while the covariance between the $\boldsymbol{x}_i$'s is assumed to be equivalent to what would be generated by the current study.[13]

[13] Using the notation in Zellner (1986), this would also involve setting $g=n$, which is termed the unit information prior.

When making use of these priors in a linear regression model, the posterior distributions would take the form:

$$\begin{aligned}\boldsymbol{\beta}|\sigma^2,\boldsymbol{y} &\sim N\left(\hat{\boldsymbol{\beta}},\frac{g}{g+1}\sigma^2\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\right) &(14)\\ \boldsymbol{\sigma}|\boldsymbol{y} &\sim IG\left(\frac{a+n}{2},\frac{1}{2}\left(b+\boldsymbol{y}^\top\boldsymbol{y}-\hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}}\left(\frac{1+1}{g}\right)\right)\right) &(15)\end{aligned}$$

where $\hat{\boldsymbol{\beta}}=\frac{g}{g+1}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{y}$, which is essentially equal to the ordinary least squares estimate, when g is large (and would usually be set to the value of n). A similar result may be derived for the variance-covariance matrix of the residual, where $p\left(\sigma^2|\boldsymbol{y}\right)\approx IG(\sigma^2;n/2,SSR/2)$. As will be shown below, this result no longer applies in those cases where we make use of model selection techniques to identify those variables that may be most useful, when p is particularly large.

To describe how the Bayesian paradigm may be used for model selection purposes in a high-dimensional setting, we could denote each possible model by $\gamma=(\gamma_1,\ldots,\gamma_p)$, where

$$\gamma_j=\begin{cases}1, & \text{if } \beta_j\neq 0\\ 0, & \text{if } \beta_j=0\end{cases} \tag{16}$$

In such a setting, the γ parameter is a random variable, and the posterior distribution $p(\gamma|\boldsymbol{y})$ could be derived for all possible variable combinations and may be used to identify the variables that should ultimately be included in the model. In terms of notation, $\boldsymbol{X}_\gamma$ and $\boldsymbol{\beta}_\gamma$ refer to the subset of $\boldsymbol{X}$ and $\boldsymbol{\beta}$ that have been selected for different values within the γ vector. The likelihood function for this model would take the form $p\left(\boldsymbol{y}|\boldsymbol{\beta}_\gamma,\sigma^2,\gamma\right)$, for which we require an associated prior for both the parameters, $p\left(\boldsymbol{\beta}_\gamma,\sigma^2|\gamma\right)$, and for all possible models, $p(\gamma)$. Using the theorem of Bayes, we could then derive the posterior distribution for the model selection and parameter estimation problem, as follows:

$$p(\gamma|\boldsymbol{y})=\frac{p(\boldsymbol{y}|\gamma)\,p(\gamma)}{p(\boldsymbol{y})}\propto p(\gamma)\int p\left(\boldsymbol{y}|\boldsymbol{\beta}_\gamma,\sigma^2,\gamma\right)p\left(\boldsymbol{\beta}_\gamma,\sigma^2|\gamma\right)d\boldsymbol{\beta}_\gamma d\sigma^2 \tag{17}$$

Note that $p(\boldsymbol{y}|\gamma)$ is the marginal likelihood that may be derived from $\int p\left(\boldsymbol{y},\boldsymbol{\beta}_\gamma,\sigma^2|\gamma\right)d\boldsymbol{\beta}_\gamma d\sigma^2$, which is the prior expected likelihood for the given model.

In this case we could make use of Zellner's prior, once again, where g defines the preferred variance for the regression coefficient, while (a,b) would usually take on small values ($a=b=0.01$), since we prefer small values for the residuals. The remaining prior, $p(\gamma)$ is also important and could be formulated to influence the number of variables that will be included in the preferred model. Hence, when we apply this framework we need to make use of the following components:

$$\begin{aligned}\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\gamma &\sim N\left(\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma,\sigma^2 I\right)\\ \boldsymbol{\beta}_\gamma,\sigma^2|\gamma &\sim N\left(0,\sigma^2 g\left(\boldsymbol{X}_\gamma^\top\boldsymbol{X}_\gamma\right)^{-1}\right)\times IG(a/2,b/2)\\ \gamma &\sim p(\gamma)\end{aligned}$$

To apply Bayesian model selection techniques to a high-dimensional setting, we could use the framework of Mitchell and Beauchamp (1988) and George and McCulloch (1993), where we employ the null and alternative hypothesis for the regression, $y_i=\beta_1 x_{i1}+\varepsilon_i$:

$$H_0: \quad \beta_1=0, \text{ with probability } P(\gamma_1=0)=0.5 \tag{18}$$

$$H_1: \quad \beta_1\sim N\left(0,g\sigma^2\right), \text{with probability } P(\gamma_1=1)=0.5 \tag{19}$$

In this case we may start off by assuming that the prior probability that β_1 should be different from zero is 0.5,

while the probability that the associated variable should be excluded is also 0.5. Hence,

$$p(\beta_1) = p(\beta_1|\gamma_1 = 0)\,P(\gamma_1 = 0) + p(\beta_1|\gamma_1 \neq 0)\,P(\gamma_1 \neq 0) \tag{20}$$

The literature includes many alternatives to the assumption that the distribution for $\boldsymbol{\beta}_\gamma$ is normal under the alternative hypothesis. For example, the use of distributions that have thick tails, such as those that are employed in Liang et al. (2008) and Bayarri et al. (2012), would potentially reduce the bias, while non-local priors may be more appropriate when dealing with sparsity. For example, Johnson and Rossell (2010) make use of a moment (MOM) prior, where under the alternative hypothesis the mass of the prior probability is not centred on zero. For additional details relating to the properties of such non-local priors, see Johnson and Rossell (2012) and Rossell and Telesca (2017).

As has been noted already, we also need to set a prior for $p(\boldsymbol{\gamma})$, which relates to the number of variables $p_\gamma = \sum_{j=1}^{p} \gamma_j$ that should be included in the model. Two prominent priors include the Beta-Binomial (1,1), which is applied in Scott and Berger (2010) and makes use of a uniform $P(p_\gamma = I) = 1/(p+1)$ distribution for all $I = 0, \ldots, p$. Such a prior is not particularly informative about the model size and implies that:

$$p(\boldsymbol{\gamma}) = \frac{1}{(p+1)\binom{p}{p_\gamma}} \tag{21}$$

The other prominent prior is termed the complexity(c) prior, which is discussed in Castillo et al. (2015) and would prefer more parsimonious models, since it assumes that $P(p_\gamma = I)$, which decreases exponentially, such that:

$$p(\boldsymbol{\gamma}) \propto \frac{1}{p^{cp_\gamma}\binom{p}{p_\gamma}} \tag{22}$$

After specifying all the necessary priors in the model we would then need to derive estimates for the model parameters. This procedure may involve one of three main strategies, where we could select the model that has the highest posterior probability, $\hat{\boldsymbol{\gamma}}$ ($\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} p(\boldsymbol{\gamma}|\boldsymbol{y})$), which would be used for the parameter estimates $E(\boldsymbol{\beta}_\gamma|\boldsymbol{y}, \hat{\boldsymbol{\gamma}})$. Such a strategy would be appropriate if $p(\hat{\boldsymbol{\gamma}}|\boldsymbol{y}) \approx 1$. However, if this is not the case then it would ignore the uncertainty that relates to the model selection process.

An alternative strategy would make use of Bayesian model averaging (BMA), which was successfully applied in Hahn and Carvalho (2015), where:

$$E(\boldsymbol{\beta}|\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}} E(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\gamma})\, p(\boldsymbol{\gamma}|\boldsymbol{y}) \tag{23}$$

$$E(y^\star|\boldsymbol{y}, \boldsymbol{x}^\star) = E(\boldsymbol{\beta}|\boldsymbol{y})^\top \boldsymbol{x}^\star \tag{24}$$

$$P(\beta_j \in [l,u]\,|\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}} P(\beta_j \in [l,u]\,|\boldsymbol{\gamma}, \boldsymbol{y})\, p(\boldsymbol{\gamma}|\boldsymbol{y}) \tag{25}$$

With the aid of these techniques we would be able to derive closed-form solutions for Zellner's prior. In other cases we may need to make use of a stochastic algorithm to approximate the likelihood function, where common variants include the Markov-Chain Monte-Carlo, Laplace, and Bayes methods.

Then lastly, the third strategy would involve making use of the posterior probabilities for the individual models, which could be used to calculate the posterior marginal probability that a specific variable should be included in the model, $p(\gamma_j = 1|\boldsymbol{y})$, where if a specific variable appears in most of the model specifications that have a high posterior probability, $p(\boldsymbol{\gamma}|\boldsymbol{y})$, then it will be assigned a large posterior marginal probability of inclusion. In such cases, we would usually report on the variables with $P(\beta_j \neq 0|\boldsymbol{y}) > s$, for some threshold, s. This would of course require that we need to set a value for s, which could utilise the median probability model, where $s = 0.5$, as in Barbieri and Berger (2004).

The literature suggests that Bayesian model selection techniques exhibit good frequentist properties when used for selection, estimation and posterior uncertainty quantification purposes. To illustrate this feature, we could consider the connection that exists between the Bayesian models and the equivalent frequentist models that

employ L_0 penalties, where the penalised regression, $h(\boldsymbol{y},\boldsymbol{\gamma}) = p\left(\boldsymbol{y}|\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}},\hat{\sigma}^2_{\boldsymbol{\gamma}}\right)e^{-\eta_{\boldsymbol{\gamma}}}$, would allow for the derivation of the following estimated parameters:

$$\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}},\hat{\sigma}^2_{\boldsymbol{\gamma}}\right) = \arg\max_{\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2} p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2\right) \tag{26}$$

If we were then to suggest that the penalty $\eta_{\boldsymbol{\gamma}}$ depends on $(p_{\boldsymbol{\gamma}},n,p)$, then:

$$\tilde{h}(\boldsymbol{y},\boldsymbol{\gamma}) = \frac{h(\boldsymbol{y},\boldsymbol{\gamma})}{\sum_{\boldsymbol{\gamma}} h(\boldsymbol{y},I)} \approx P(\boldsymbol{\gamma}|\boldsymbol{y}) \tag{27}$$

Now if $p\left(\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2|\boldsymbol{\gamma}\right) = N\left(0,g\sigma^2\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\right)\times IG$, then $p(\boldsymbol{\gamma}) = (1+g)^{p_{\boldsymbol{\gamma}}}/\eta_{\boldsymbol{\gamma}}$. This derivation facilitates a direct link between the frequentist models that employ L_0 penalised likelihood functions and Bayesian model selection techniques. For example, when $\eta_{\boldsymbol{\gamma}}$ is set to $0.5p_{\boldsymbol{\gamma}}\log(n)$, then we would be making use of the L_0 BIC penalty, which may be approximated with a high degree of accuracy, by making use of Bayesian model selection techniques, where the priors are set to $g = n$ and $p(\boldsymbol{\gamma}) = 1/2^p$. Similarly, the extended BIC penalty of Chen and Chen (2008), which is used in this study, involves employing the L_0 penalty, $0.5p_{\boldsymbol{\gamma}}\log(n) + \log\binom{p}{p_{\boldsymbol{\gamma}}}$, which is equivalent to setting $g = n$ and $p(\boldsymbol{\gamma}) = \text{Beta-Binomial}(1,1)$. For additional details relating to the theoretical results of these models, see Rossell (2021) and Rousseau and Szabo (2020).

In our case we make use of Bayesian model selection techniques to identify the single model that is responsible for highest marginal posterior probability. In addition, we also make use of Bayesian model averaging techniques to select the predictors that are included in models that have a marginal posterior probability that is greater than 0.1.

2.3 Nonlinear statistical learning models

2.3.1 Ensemble methods

For comparative purposes, we have also made use of an ensemble method, which takes the form of the complete subset regression (CSR) framework of Elliott et al. (2013, 2015). This procedure makes use of the results from independent models that are then combined with a deterministic calculation and is in many respects similar to the bagging procedure of Breiman (1996). It provides an intuitive method for generating forecasts from many variables. To apply this methodology we fit a linear regression model that seeks to explain y_i using each of the individual regressors in x_{i-h}. To identify the “best" predictors, we would then rank the absolute value of the t-statistics from the initial coefficient estimates. These predictors are then used to generate a number of individual forecasts, which are combined to provide the CSR forecast.

2.3.2 Random forests

The random forest model of Breiman (2001) reduces the variance of regression trees, which are nonparametric models that approximate an unknown nonlinear function with local predictions using recursive partitioning of the parameter space. They are based on bootstrap aggregation (bagging) methods for randomly constructed regression trees that take the form of nonparametric models that approximate an unknown nonlinear function with local predictions, using recursive partitioning of the parameter space that pertains to the covariates.

For a given number of terminal nodes, K, the parameter space would be split to minimise the sum of squared errors in the regression:

$$y_{i+h} \quad = \quad \sum_{k=1}^{K} c_k I_k(\boldsymbol{x}_i;\boldsymbol{\theta}_k) \tag{28}$$

where $I_k(x_i;\theta_k)$ is an indicator function such that,

$$I_k(x_i;\theta_k) = \begin{cases} 1 & \text{if } x_i \in R_k(\theta_k), \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

2.3.3 Gradient boosting

As an alternative to random forests, gradient boosting seeks to build a model by repeatedly fitting a regression tree to the residuals. After each tree has grown to model the residuals, it is shrunk down by a factor before it is added to the current model. This would allow us to explain certain elements (including non-linear relationships) that may have been discarded in the residual. A general gradient descent “boosting" paradigm has been developed for additive expansions based on any fitting criterion. It utilises developments discussed in Friedman et al. (2000) and Friedman (2001), where special enhancements are derived for the particular case where the individual additive components are regression trees. In general, it has been suggested that gradient boosting of regression trees produces competitive, highly robust results.

2.3.4 Deep learning (neural networks)

Neural network models usually take the form of highly parameterised non-parametric specifications that are able to potentially explain any non-linear function. These models would often make use of a large number of parameter weights that transform the data that is contained in the set of predictors to fit the target variable. These parameter weights are learnt through repeated exposure to different subsets of the data. Deep learning methods make use of layered representations of neural networks models that are stacked on top of each other to provide a mathematical framework for learning the rules that would allow for the mapping of predictors to the target variable. Such models could potentially explain behaviour that is extremely complex, although there is also a significant possibility that the model may be prone to over-fitting errors. In our case, we have made use of a relatively simple model structure that will hopefully circumvent such concerns, where we have incorporated three hidden layers and a relatively parsimonious combination of 32, 16, and 8 nodes.

3 Data

The South African Consumer Price Index (CPI) measures changes in the general level of prices of consumer goods and services. It is a fixed-basket price index, in that it represents the cost of purchasing a fixed basket of consumer goods and services of constant quality and similar characteristics (Statistics South Africa, 2017a). The items that are included in the basket seek to represent average household expenditure, using information from the Income and Expenditure Survey (IES) and more recently from the Living Conditions Survey (LCS), which was last conducted in 2014/15.[14] Note that the index only incorporates data on those products that contribute at least 0.1% of the total household expenditure. Additional data sources such as regulatory reports, excise tax receipts, industry association reports and summarised transaction data from retailers, are then used to align the data from the respective surveys with the data that goes into the household final consumption expenditure in the national accounts. The last update to the items that are included in the CPI basket was in January 2017 and the next update is expected to take place during 2021 (Statistics South Africa, 2017b).

Since 2006, StatsSA has made use of fieldworkers who are responsible for collecting the relevant prices from the retail outlets directly. Each province has its own basket and every product that appears in at least one provincial basket is included in the national basket. The current CPI contains 412 products, which is slightly more than the previous basket, which included 393 products (Statistics South Africa, 2017b) and its composition follows the United Nations Statistical Division (UNSD) standard for classifying household expenditure on goods and services. This standard is termed the Classification of Individual Consumption by Purpose (COICOP) and it currently incorporates 14 high-level (or 2-digit) categories (01-Food and non-alcoholic beverages). Table 1, which

[14] The LCS surveys household expenditure in precisely the same way as the IES, but also measures a range of additional poverty indicators.

is taken from Statistics South Africa (2017a), shows how the naming convention of the COICOP has been applied to the different levels of products and categories.

COICOP	Level	Name	Example
01	2-digit	Category	Food and non-alcoholic beverages
011	3-digit	Class	Food
0111	4-digit	Group	Bread and cereals
01112	5-digit	Product	Bread
01112001	8-digit	Commodity	Loaf of white bread
01112001*wxyz*	12-digit	Sampled product	Loaf of white bread for specific brand, size, outlet (within an area)

Table 1: Convention for COICOP classification

In the subsequent analysis, we make use of the monthly four-digit data on consumer prices between January 2008 to March 2021, since a slightly different methodology was used to collect and classify the data for prior periods of time.[15] In addition, we have also made use of a new dataset that contains more disaggregated data on the prices of goods in the consumption basket. This dataset includes information on 216 products or categories, where food products are measured at the 8-digit level and all other goods and services are measured at the 5-digit level. Unfortunately, the first observation in this dataset relates to January 2017, which would make for an extremely small in-sample training period in our case. Therefore, with the help of StatsSA we have now extended this dataset, going back to January 2009, by making use of the fieldworker data, which has been collected for 34,075 different products, across 5,505 outlets, that arise in 85 different areas.

To obtain a measure for the changes in prices over time, we calculate the price relative indices on the available fieldworker data, utilising the method that is used in the compilation of the respective CPI indices. This procedure involves the construction of a Jevon's index, which is defined as the unweighted geometric mean of the price ratios that utilise data for the current and previous survey periods for a particular commodity (i.e. at the 8-digit level). Such a Jevon's index may be constructed as follows,

$$I_i^J = \prod_{i=1}^{n} \left(\frac{P_{\theta,i}}{P_{\theta,i-1}} \right)^{1/\xi} \tag{30}$$

where I_t^J denotes the Jevon's index, while $P_{\theta,t}$ is the price of commodity θ in period i, and ξ refers to the total number of items that are included in this calculation. In this study we calculate a number of different variants of price relative indices to obtain information about price movements. This is then used to construct individual indices for each of the components at different levels of aggregation.

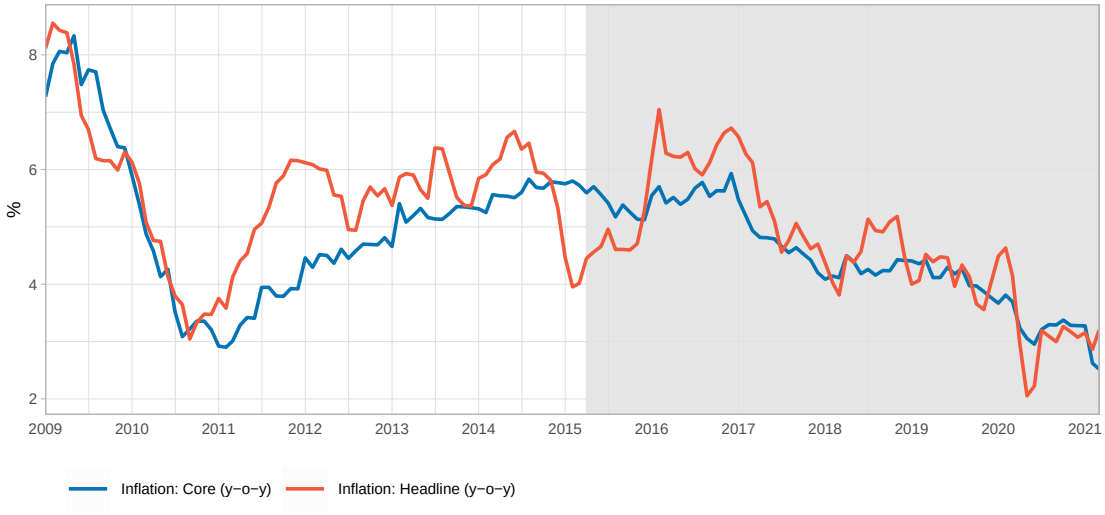
Figure 1 displays the measures of headline inflation and core inflation over the entire sample, where the shaded area relates to all the time periods for which we assume that we do not have information on the predictors, when estimating the parameters in a statistical learning model, that is subsequently used to generate a 24-month-ahead forecast. It illustrates that the trend in both measures of inflation have declined over the out-of-sample period, which would suggest that most mean reverting models will produce a negative forecast bias. In addition, as would be expected, headline inflation is certainly much more volatile than core inflation, where over the out-of-sample period core inflation has a variance of 0.61%, while the variance of headline inflation is 1.04%.

4 Results

To evaluate the relative performance of the different models we make use of a recursive out-of-sample forecasting exercise and a forecasting horizon of between one and 24 months ahead. The statistics that are used to

[15] These changes are discussed in Statistics South Africa (2007).

Figure 1: Inflation over initial in-sample and out-of-sample period (year-on-year)



evaluate the out-of-sample performance of the respective models include the root-mean squared-error (RMSE), the mean absolute percentage error (MAPE) and the Diebold and Mariano (1995) statistics.[16] When reporting on the results, we compare the year-on-year inflation forecasts against the official CPI release for headline and core (i.e. excluding food and non-alcoholic beverages, fuel and energy) inflation. To generate forecasts we mostly make use of the direct forecasting approach, where the only exceptions pertain to the random walk, DIM, autoregressive, and vector autoregressive models.[17] To apply the direct forecast approach over the forecasting horizon of $h = \{1,\dots,24\}$, we estimate the model $y_i = \sum_{j=1}^{p} x_{i-h,j}\beta_j + \varepsilon_i$, and use the coefficients to find $E_i\left[y_{i+h}|x_{i,j}\right] = \sum_{j=1}^{p} x_{i,j}\hat{\beta}_j$, where the predictors may include lagged values of the target variable. In addition, for comparative purposes we also make use of the conditional forecasting function that was employed in Medeiros et al. (2021), which may be expressed as $E_i\left[y_{i+h}|\left(y_{i+h-1},x_{i,j}\right)\right]$. These two expressions would provide equivalent forecasts, when $h = 1$, however, by way of example, when the forecast horizon is 12 steps ahead, the conditional forecasting function would make use of the value of inflation that will be observed 11 months into the future to estimate a coefficient that is then used to generate the 12-month-ahead forecast. In contrast, the direct forecasting approach does not use any future observations to generate forecasts for inflation.

In addition, to these statistics we also follow Lundberg and Lee (2017) and Joseph et al. (2020) and compute Shapley values for the high-dimensional regression model that appears to provide the most attractive out-of-sample results. Shapley (1953) derived a method for assigning payoffs to players depending on their contribution to the total payoff, where players cooperate in a coalition and receive a portion of the payoff from this cooperation. This concept may be applied to enhance the interpretability of the results from any machine learning model, where the "game" is the task of prediction and the "gain" is the prediction that is provided after including the regressor minus the average prediction that was generated when incorporating all the regressors. Hence, the Shapley value is the average marginal contribution of the selected regressors to the prediction, where the value of the j-th regressor would contribute ϕ_j to the prediction compared to the average prediction from all the regressors.

From an intuitive perspective the Shapley value represents the average change in the prediction that is attributed to the regressor after it is incorporated into the model. To measure the marginal contribution of a regressor to the prediction, we need to consider all the possible combinations of the regressors:

[16] Although there will be cases where these models will be nested, which would imply that the use of the statistics that are discussed in Clark and West (2007) and McCracken (2007) would be preferred to the Diebold and Mariano (1995) statistic, these models are not always nested within one another. Therefore, for the sake of consistency, we make use of the Diebold and Mariano (1995) statistic to evaluate all of the models. This would suggest that the results may favour the more parsimonious model in those cases where the models are nested.

[17] Explicit forecast functions for these models have been included in the above description of the respective models.

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right) \tag{31}$$

where S is a subset of the regressors used in the model, x is the vector of coefficient values and p refers to the number of features. The prediction from the regressors in set S, which is presented by $val_x(S)$, are marginalised over the regressors that are not included in set S, such that:

$$val_x(S) = \int \hat{f}(x_1,\ldots,x_p) dP_{x \notin S} - E_X(\hat{f}(X)) \tag{32}$$

Therefore, to compute these statistics we would need to perform multiple integrations for each regressor that is not contained in S.

4.1 Four-digit data: headline inflation

Headline inflation is made up of 46 different price indices that are measured at the four-digit level of aggregation. These indices are used to generate the DIM forecast, which has a significant influence over the official central bank near-term monthly inflation forecast. Given the relatively small number of predictors, there are sufficient degrees-of-freedom to estimate a linear regression model. Table 2 contains the unconditional out-of-sample RMSE statistics, when we make use of the direct forecasting approach. When comparing all the benchmark models, we note that with the exception of the linear regression model, the errors are all fairly similar, where the DIM and official SARB forecasts are superior over the short-term, while the random walk and stochastic volatility models are superior over longer horizons. Note also that over the first three months, the official SARB forecasts provide a RMSE that is about half the size of the DIM, which suggests that the use of "off-model" information has reduced the forecasting error by a relatively large amount over these horizons.

Turning our attention to the relative forecasting performance of the linear regression model, we note that it provides results that are indicative of a model that is prone to the over-fitting problem, as those models that make use of variable selection techniques would usually provide more impressive results. This would also suggest that the matrix that contains the predictors may be sparse, although we do not make use of a specific definition for statistical sparsity, as in McCullagh and Polson (2018). Further support for this finding is included in appendix C, where the in-sample estimation results for the full sample suggest that a model that makes use of 12 explanatory variables is able to provide a near perfect explanation of the behaviour of headline inflation.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.41	0.68	0.82	0.93	1.09	1.68	1.75	1.68
AR-DIRECT	0.42	0.7	0.85	0.97	1.12	1.55	1.66	1.69
STOCH-VOL	0.42	0.64	0.76	0.84	0.91	1.21	**1.31**	**1.38**
RAND-WALK	0.41	0.64	0.76	0.84	0.9	**1.19**	1.33	1.45
BVAR	0.6	0.79	0.93	1.04	1.2	1.41	1.45	1.47
DIM	0.36	0.59	0.72	0.82	0.97	1.49	1.61	1.69
SARB	**0.15**	**0.25**	**0.37**	**0.57**	**0.84**	1.42	1.59	1.63
LINEAR	0.5	0.8	1.15	1.11	1.56	2.42	3.06	3.05
DFM-TF	0.43	0.68	0.84	0.97	1.2	1.85	1.98	1.89
DFM-3PRF	0.45	0.7	0.85	0.97	1.23	1.8	1.77	2.03
LASSO	0.39	0.73	0.94	1.16	1.56	2.11	2.91	2.28
LASSO-PSI	0.76	1.87	1.49	1.25	1.82	3.4	3.25	2.53
ADAP-LASSO	0.4	0.74	1	1.18	1.6	2.14	2.74	2.16
POST-LASSO	0.44	0.68	0.66	0.76	1.09	1.77	1.78	2.53
LASSO-ZERO	0.42	0.7	0.87	1.04	1.2	2.46	2.19	3.47
RIDGE	0.48	0.9	1.13	1.29	1.68	2.25	3.07	2.46
ELASTIC	0.39	0.79	1.02	1.18	1.68	2.11	3.04	2.37
ADAP-ELASTIC	0.39	0.81	1.07	1.23	1.7	2.19	2.97	2.31
SCAD	0.44	0.83	0.88	1.02	1.43	2.2	2.12	3.33
BMS	0.44	0.71	0.9	0.95	1.13	2.43	1.5	2.88
BMA	0.44	0.72	0.78	0.87	0.94	1.34	1.4	1.68
CSR	0.41	0.65	0.75	0.82	0.92	1.77	1.66	2.11
NEURAL-NET	0.45	0.74	0.96	1.01	1.21	2	1.83	1.85
RANDOM-FOREST	0.51	0.72	0.9	1.03	1.25	1.43	1.57	1.64
XG-BOOST	0.45	0.64	0.78	0.95	1.1	1.28	1.33	1.48

Table 2: Unconditional root-mean squared-error (four-digit headline inflation)

Model acronyms: “AR(1)" - first-order autoregressive, “AR-DIRECT" - first-order autoregressive with direct forecasting function, “STOCH-VOL" - stochastic volatility, “RAND-WALK" - random-walk, “BVAR" - large Bayesian vector autoregressive model, “DIM" - disaggregated inflation model, “SARB" - official SARB forecast reported to MPC, “LINEAR" - lienar regression model, “DFM-TF" - dynamic factor model with target factors, “DFM-3PRF" - dynamic factor model with three pas filter, “LASSO" - least absolute shrinkage and selection operator, “LASSO-PSI" - LASSO with post-selection inference, “ADAP-LASSO" - adaptive LASSO, “POST-LASSO" - post-OLS LASSO, “LASSO-ZERO" - LASSO with L_0 penalty, “RIDGE" - ridge regression, “ELASTIC" - elastic net, “ADAP-ELASTIC" - adaptive elastic net, “SCAD" - smoothly clipped absolute deviation, “BMS" - Bayesian model selection, “BMA" - Bayesian model averaging, “CSR" - complete subset regression, “NEURAL-NET" - neural network, “RANDOM-FOREST" - random forest, “XG-BOOST" - boosting.

When we compare the accuracy of the forecasts from the “dense” models relative to the “sparse” models we note that the results are somewhat mixed, where although there are a number of cases where the variable selection techniques provide more impressive results, the DFMs are at least competitive in all cases and superior over longer horizons. In the case of the 24-month-ahead forecast, this would imply that the observed values of the predictors from 24 months ago, which provide the best explanation of current headline inflation are not necessarily going to be the same, as the ones that provide the best 24-month-ahead forecast, from the current point in time. Furthermore, we also note that in this case, the variable selection techniques would in most cases appear to be inferior to the benchmarks, which include the autoregressive, stochastic volatility and random walk models. The results for the nonlinear statistical learning models are in most cases similar to those of the DFMs models, however, over horizons that are longer than six months, the model that makes use of boosting methods provides

forecasts that are more accurate than both “dense” and “sparse” models.

These results differ to what is produced when we make use of a conditional forecasting function. The results for which are contained in appendix D.1, where the performance of the statistical learning models is superior to what is provided by the respective benchmarks. In addition, when making use of a conditional forecasting function the “sparse” models are responsible for more accurate forecasts, relative to the “dense” models, while over longer horizons the neural network and boosting models provide forecasts that are slightly more accurate than the “sparse” models. These results also suggest that the neural network provides the lowest RMSE at the 4-, 6-, 12- and 24-step-ahead horizon. Similar results are provided by the MAPE statistics, which are contained in appendix E.1.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.06	-0.84	-0.72	-0.82	-0.88	-0.68	-0.45	-0.32
AR-DIRECT	-0.2	-1.65	-1.21	-1.26	-1.41	-0.79	-0.5	-0.38
STOCH-VOL	-0.26	0.04	0.1	0.07	-0.12	-0.07	0.04	0.17
BVAR	**-2.21**	-1.77	-1.94	-1.6	-1.16	-0.46	-0.19	-0.03
DIM	*2.19*	1.46	0.91	0.46	-0.61	-0.78	-0.5	-0.43
SARB	*3.67*	*2.3*	1.84	1.88	0.35	-0.61	-0.56	-0.37
LINEAR	-1.46	-1.94	**-2.3**	-1.84	**-1.97**	-1.26	-1.33	**-4.22**
DFM-TF	-0.77	-0.71	-1.05	-1.21	-1.19	-1.09	-0.62	-0.44
DFM-3PRF	-1.36	-1.04	-1.44	-1.39	-1.44	-0.89	-0.48	-1.21
LASSO	0.51	-0.89	-1.41	**-2.16**	**-3.22**	-1.66	-1.04	**-3.54**
LASSO-PSI	**-2.58**	**-2.4**	**-2.94**	**-2.72**	**-3.96**	**-3.04**	-1.26	-1.81
ADAP-LASSO	0.43	-0.94	**-2.08**	**-2.3**	**-2.77**	-1.63	-0.86	**-3.09**
POST-LASSO	-0.97	-0.68	1.05	0.81	-0.78	-1.15	-0.75	**-2.19**
LASSO-ZERO	-0.32	-1.68	-1.01	-1.76	-0.89	**-7.11**	**-2.73**	-0.97
RIDGE	-0.99	**-2.01**	**-2.26**	**-2.46**	**-4.28**	**-2.93**	**-5.03**	**-4.61**
ELASTIC	0.5	-1.34	**-2.31**	**-2.23**	**-3.75**	-1.38	-1.15	**-4.17**
ADAP-ELASTIC	0.53	-1.45	**-2.52**	**-2.62**	**-3.32**	-1.27	-1.13	**-4.07**
SCAD	-1.01	**-2.46**	-1.41	**-2.54**	-1.62	**-2**	-1	**-3.48**
BMS	-1.5	-1.62	-1.41	-0.97	-0.72	**-5.15**	-0.48	**-2.85**
BMA	-0.95	-1.33	-0.15	-0.31	-0.33	-0.82	-0.55	**-3.77**
CSR	0.43	-0.17	0.21	0.4	-0.11	-0.9	-0.38	**-4.61**
NEURAL-NET	-1.05	-1.39	**-2.13**	**-2.63**	-1.59	-0.97	-0.45	-0.82
RANDOM-FOREST	-1.24	-1.05	-1.53	-1.65	-1.21	-0.5	-0.42	-0.4
XG-BOOST	-1.37	0.03	-0.28	-1.78	-1.07	-0.21	0	-0.24

Table 3: Diebold-Mariano statistics (four-digit head)
Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

Table 3 contains the Diebold and Mariano (1995) statistics for the forecasts that use the unconditional forecasting function, relative to the forecasts from the random-walk model. In this case we note that the only forecasts that are significantly superior to the random-walk forecast are provided by the DIM over a one-month horizon and by the SARB forecast over a one- and two-month horizon. Furthermore, the forecasting performance of the DFMs relative to the random-walk forecast is not significantly different from zero over all horizons, while the random-walk forecast provides a significant improvement in forecasting performance over several horizons, relative to most of

the models that employ variable selection techniques.

4.2 Eight/five-digit data: headline inflation

When using data that is measured at the eight-digit level of aggregation for food items and at the five-digit level for most other goods, we have a total of two-hundred and ten different price indices for headline inflation. Given the relatively large number of predictors, we do not have sufficient degrees-of-freedom to estimate a linear regression model. Table 4 contains the out-of-sample RMSEs for the different models, where most of the results that pertain to the benchmarks are similar to what was provided when using four-digit data, with the exception of the Bayesian VAR, which have deteriorated slightly.

Note that the RMSEs for "sparse" models are lower when using the more disaggregated data, which would suggest that the combined use of more disaggregated data and variable selection techniques allows for an improved forecasting performance, as it would discard some of the noise that may be included in the variables when they are subject to greater degrees of aggregation. This is in contrast with the results of the "dense" models which provide forecasts that are slightly more inaccurate than those that were derived from the four-digit data. And then finally, the results for the nonlinear statistical learning models are in most cases comparable to those that make use of the four-digit data.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.41	0.68	0.83	0.95	1.13	1.72	1.76	1.66
AR-DIRECT	0.41	0.69	0.83	0.94	1.08	1.41	1.47	1.49
STOCH-VOL	0.42	0.65	0.76	0.84	0.92	1.23	1.34	**1.41**
RAND-WALK	0.41	0.64	0.76	0.84	0.9	**1.19**	**1.33**	1.45
BVAR	0.51	0.68	0.79	0.9	1.09	1.45	1.61	1.72
DIM	0.36	0.59	0.72	0.82	0.97	1.49	1.61	1.69
SARB	**0.15**	**0.25**	**0.37**	**0.57**	**0.84**	1.42	1.59	1.63
DFM-TF	0.43	0.66	0.82	0.98	1.28	1.89	1.69	1.74
DFM-3PRF	0.46	0.7	0.82	0.95	1.3	2.03	1.97	2.38
LASSO	0.43	0.62	0.65	0.8	1.18	1.85	1.79	1.85
LASSO-PSI	1.59	1.18	2.02	1.98	1.66	2.51	2.33	2.08
ADAP-LASSO	0.43	0.62	0.66	0.81	1.18	1.86	1.79	1.9
POST-LASSO	0.46	0.8	0.91	1.07	1.35	1.69	1.94	1.99
LASSO-ZERO	0.53	1.51	2.21	2.62	5.75	3.03	2.24	2.02
RIDGE	0.37	0.57	0.72	0.81	1.2	2.24	1.8	2.55
ELASTIC	0.42	0.6	0.66	0.78	1.13	1.72	1.84	2
ADAP-ELASTIC	0.43	0.58	0.67	0.77	1.15	1.72	1.91	2
SCAD	0.41	0.78	0.93	1.01	1.44	2.09	1.86	2.33
BMS	1.57	2.26	2.18	2.41	2.31	3.65	3.06	2.39
BMA	2.24	1.85	2.45	2.77	2.53	4.74	2.98	2.8
CSR	0.41	0.66	0.84	1.02	1.21	1.63	1.91	2.22
NEURAL-NET	0.5	0.66	0.8	0.99	1.06	1.48	1.5	1.45
RANDOM-FOREST	0.58	0.8	1.02	1.2	1.32	1.52	1.7	1.7
XG-BOOST	0.49	0.74	0.99	1.15	1.39	1.57	1.71	1.72

Table 4: Unconditional root-mean squared-error (8/5-digit headline)
See Table 2 for model acronyms.

Once again, the results from the unconditional forecasting function differ to those that employ the conditional forecasting function, which are contained in appendix D.2, where we note that it is suggested that the nonlinear statistical learning models provide superior forecasts to the models that employ variable selection techniques, when the horizon is greater than three months ahead. Furthermore, in that case, the models that employ variable selection techniques are superior to both the benchmarks and the DFMs. Once again, similar results are evident from the MAPE statistics, which are contained in the appendix.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.11	-0.84	-0.8	-0.98	-1.09	-0.8	-0.55	-0.31
AR-DIRECT	0.02	-1.06	-0.86	-0.9	-0.9	-0.46	-0.21	-0.06
STOCH-VOL	-0.37	-0.1	-0.03	-0.13	-0.27	-0.15	-0.03	0.1
BVAR	-1.42	-0.53	-0.54	-0.69	-1.04	-0.58	-0.39	-0.22
DIM	*2.19*	1.46	0.91	0.46	-0.61	-0.78	-0.5	-0.43
SARB	*3.67*	*2.3*	1.84	1.88	0.35	-0.61	-0.56	-0.37
DFM-TF	-0.69	-0.29	-0.7	-1.33	-1.42	-1.31	-0.47	-0.58
DFM-3PRF	-1.32	-0.85	-0.67	-0.78	-1.62	-0.87	-0.81	**-2.84**
LASSO	-0.37	0.51	*2.09*	0.55	-1.03	-1.34	-1.05	-1
ADAP-LASSO	-0.37	0.56	1.8	0.42	-1.03	-1.35	-1.05	-1.5
POST-LASSO	-1.38	-1.55	**-1.98**	**-2.64**	-1.29	-0.75	-0.77	-1.7
LASSO-ZERO	**-2.94**	**-3.04**	**-3.95**	**-2.81**	-1.59	**-4.79**	-1.83	-0.56
RIDGE	0.92	1.08	0.53	0.29	-0.83	-0.92	-0.73	**-3.42**
ELASTIC	-0.19	0.68	1.54	0.81	-0.83	-0.94	-1.12	-1.8
ADAP-ELASTIC	-0.28	1.06	1.4	0.91	-0.89	-0.93	-1.37	**-2.14**
SCAD	0.25	-1.67	-1.81	**-2.47**	-1.67	-0.97	-0.94	-1.1
BMS	**-3.71**	**-4.83**	**-4.69**	**-2.93**	**-3.3**	**-3.69**	**-2.24**	-0.79
BMA	**-4.09**	**-4.14**	**-5.68**	**-3.06**	**-3.78**	**-5.77**	**-2**	-0.99
CSR	0.15	-0.38	-1.36	**-2.03**	-1.42	-0.7	-0.56	-1.41
NEURAL-NET	-1.95	-0.49	-0.72	-1.61	-1.5	-0.64	-0.5	0.01
RANDOM-FOREST	-1.93	**-2.26**	**-3.25**	**-2.82**	-1.66	-0.6	-0.44	-0.38
XG-BOOST	**-2.17**	-1.43	**-3.16**	**-2.18**	-1.86	-0.64	-0.47	-0.47

Table 5: Diebold-Mariano statistics (8/5-digit head)

Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

Table 5 contains the Diebold and Mariano (1995) statistics, which are measured relative to the forecasts from the random-walk model. Once again, the only forecasts that are significantly superior to the random walk are provide by the DIM over a one-month horizon and by the SARB forecast over a one- and two-month horizon. Furthermore, the forecasting performance of the LASSO at the three-month horizon is significantly more accurate than what is provided by the random-walk model, while most of the other variable selection forecasts are either positive (which is due to their lower RMSE) or not significantly different from zero.

4.3 Four-digit data: core inflation

In what follows, we repeat the above analysis, but in this case, the target variable is core inflation, which is derived from the measure of CPI that excludes the effects of food, non-alcoholic beverages, fuel and energy. When using the four-digit data, the dataset for core inflation includes thirty-three different predictors. After applying this data

to the respective models we calculate the RMSEs, which are displayed in Table 6. Note that in this case, the SARB forecasts are superior over a one- and two-month horizon, while the random-walk model provides superior forecasts for between three and six months ahead. Thereafter, the lowest RMSEs are provided by the ridge and random forest models. Once again, some of the worst results are provided by the linear regression model and after generating the in-sample summary statistics for the models that make use of variable selection techniques, we observe that the matrix that contains the predictors displays sparse characteristics.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.19	0.3	0.39	0.48	0.58	0.97	1.17	1.26
AR-DIRECT	0.19	0.3	0.38	0.48	0.59	0.99	1.2	1.29
STOCH-VOL	0.18	0.28	0.34	0.41	0.51	0.83	1.01	1.11
RAND-WALK	0.18	0.26	**0.32**	**0.39**	**0.47**	0.76	0.94	1.04
BVAR	0.33	0.47	0.57	0.65	0.78	0.99	1.04	1.05
DIM	0.2	0.33	0.42	0.52	0.68	1.22	1.36	1.45
SARB	**0.13**	**0.24**	0.35	0.46	0.63	1.19	1.25	1.32
LINEAR	0.24	0.4	0.5	0.66	0.69	2.36	1.76	1.65
DFM-TF	0.23	0.36	0.48	0.62	0.82	1.44	1.6	1.8
DFM-3PRF	0.24	0.35	0.46	0.63	0.95	1.55	1.44	1.43
LASSO	0.28	0.37	0.43	0.49	0.67	1.38	1.28	1.27
LASSO-PSI	0.47	0.56	0.62	1	0.64	1.33	1.18	1.84
ADAP-LASSO	0.27	0.38	0.44	0.5	0.69	1.4	1.27	1.26
POST-LASSO	0.25	0.33	0.44	0.54	0.76	1.34	1.04	1.53
LASSO-ZERO	0.18	0.29	0.39	0.54	0.65	1.52	1.46	6.18
RIDGE	0.26	0.41	0.61	0.75	0.71	1.76	1.34	**0.96**
ELASTIC	0.25	0.37	0.45	0.49	0.61	1.39	1.23	1.22
ADAP-ELASTIC	0.26	0.39	0.46	0.49	0.63	1.4	1.2	1.25
SCAD	0.18	0.29	0.33	0.39	0.68	1.02	1.22	1.51
BMS	0.23	0.38	0.48	0.59	0.64	1.93	1.41	5.28
BMA	3.11	1.81	2.27	2.84	1.96	2.8	2.62	2.69
CSR	0.2	0.31	0.41	0.51	0.71	1.31	1.37	1.38
NEURAL-NET	0.3	0.44	0.47	0.56	0.74	1.2	1.31	1.25
RANDOM-FOREST	0.27	0.41	0.52	0.63	0.8	**0.71**	**0.85**	1.17
XG-BOOST	0.23	0.37	0.46	0.58	0.73	0.84	0.88	1.28

Table 6: Unconditional root-mean squared-error (four-digit core)
See Table 2 for model acronyms.

These results differ to those that make use of the conditional forecasting function, which are reported in appendix D.3, where the results from the "sparse" models are superior to those of the "dense" models in most cases. In addition, for most of the sparse models, the results are very close to the benchmark models over the shorter horizons, while the forecasting errors are about half the size of the benchmark models over longer horizons. Then lastly, when considering the results for the nonlinear statistical learning models we note, once again, that the neural network and boosting models provide more accurate forecasts for horizons that exceed three months. Similar results are provided by the MAPE statistics, which are contained in appendix E.3.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	**-2.5**	**-2.25**	**-2.29**	-1.89	-1.16	-0.56	-0.39	-0.33
AR-DIRECT	**-2.62**	**-2.65**	**-2.55**	**-2.79**	**-2.05**	-0.98	-0.67	-0.49
STOCH-VOL	**-2.57**	**-2.44**	**-2.39**	-1.95	-1.72	-0.92	-0.51	-0.41
BVAR	**-3.96**	**-3.02**	**-2.77**	**-2.28**	-1.72	-0.65	-0.22	-0.03
DIM	-1.78	**-2.6**	**-2.91**	**-3.02**	**-3.33**	**-2.21**	-1.25	-0.89
SARB	1.37	0.49	-0.68	-1.49	-1.82	-1.84	-0.84	-0.56
LINEAR	**-2.87**	**-2.55**	**-2.84**	**-2.86**	**-4.2**	-1.5	**-2.42**	-0.72
DFM-TF	**-3.59**	**-2.89**	**-2.49**	**-2.31**	**-2.27**	**-2.87**	-1.54	-1.77
DFM-3PRF	**-3.15**	**-2.46**	**-2**	**-2.06**	**-1.98**	**-4.22**	-0.94	-0.69
LASSO	**-3.64**	**-2.21**	**-2.12**	**-2.08**	-1.68	**-5.54**	**-2.42**	-1.35
LASSO-PSI	**-5.24**	**-6.47**	**-4.29**	**-3.67**	**-2.08**	**-2.03**	-1.45	-0.79
ADAP-LASSO	**-3.98**	**-2.27**	**-2.47**	**-2.23**	-1.88	**-5.72**	**-4.57**	**-2**
POST-LASSO	**-2.9**	-1.55	-1.83	-1.88	-1.29	**-3.55**	-0.37	-1.49
LASSO-ZERO	-0.96	-0.89	-1.23	**-2.2**	**-2.33**	**-11.19**	-1.26	-0.67
RIDGE	**-2.84**	**-2.63**	**-2.3**	-1.44	**-2.89**	-1.93	**-2.33**	0.56
ELASTIC	**-3.09**	**-2.1**	**-2.58**	**-2.17**	-1.55	**-3**	**-6.79**	**-4.93**
ADAP-ELASTIC	**-3.81**	**-2.25**	**-2.53**	**-2.09**	-1.61	**-2.36**	**-4.82**	**-5.11**
SCAD	0	-1.47	-0.49	-0.09	-1.78	-1.7	-1.36	-0.9
BMS	**-2.74**	**-2.19**	**-2.71**	**-2.86**	-1.87	-1.34	**-2.28**	-0.68
BMA	**-7.41**	**-4.59**	**-5.32**	**-2.27**	**-3.52**	**-3.09**	-1.58	-1.88
CSR	**-2.37**	-1.86	**-2.76**	**-4.31**	**-4.59**	-1.78	-0.76	-0.55
NEURAL-NET	**-3.87**	**-3.07**	**-2.11**	-1.81	**-2.32**	-0.89	-0.49	-0.31
RANDOM-FOREST	**-3.07**	**-2.67**	**-3.1**	**-3.59**	**-3.24**	0.29	0.36	-0.32
XG-BOOST	**-2.72**	**-2.16**	**-2.01**	**-2.46**	**-3.42**	-0.66	0.15	-0.5

Table 7: Diebold-Mariano statistics (four-digit core)
Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

Table 7 contains the Diebold and Mariano (1995) statistics, which suggest that there is no occasion where the difference in forecasting performance, relative to the random walk, is significantly different from zero, in favour of the competing model (even in the case of the short-term SARB forecasts). In addition, there are also a number of occasions where the random-walk model provides results that are significantly more accurate than any of the competing models.

4.4 Eight/five-digit data: core inflation

After excluding those items that are not included in the definition of core inflation, we are left with eighty-three price indices, which are measured at a five-digit level since it does not include any food items. Table 8 contains the unconditional out-of-sample RMSEs for the different models, where we note that the results are fairly similar to the case where we made use of less disaggregated data. In this case, there is only one occasion where the random-walk model does not generate the lowest RMSE over the medium to long horizon.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.18	0.28	0.36	0.45	0.55	1	1.27	1.43
AR-DIRECT	0.19	0.29	0.36	0.45	0.55	0.89	1.06	1.12
STOCH-VOL	0.18	0.27	0.34	0.41	0.5	0.8	0.98	1.07
RAND-WALK	0.18	0.26	**0.32**	**0.39**	**0.47**	**0.76**	0.94	**1.04**
BVAR	0.3	0.42	0.51	0.59	0.72	1.01	1.13	1.22
DIM	0.2	0.33	0.42	0.52	0.68	1.22	1.36	1.45
SARB	**0.13**	**0.24**	0.35	0.46	0.63	1.19	1.25	1.32
DFM-TF	0.18	0.27	0.33	0.39	0.55	1.09	1.31	1.46
DFM-3PRF	0.27	0.43	0.6	0.77	1.08	1.63	1.47	1.57
LASSO	0.28	0.36	0.38	0.55	0.83	1.49	1.04	1.51
LASSO-PSI	0.52	0.85	0.96	0.65	1.03	1.09	1.72	1.37
ADAP-LASSO	0.27	0.37	0.37	0.52	0.82	1.52	1.03	1.52
POST-LASSO	0.22	0.36	0.45	0.54	0.72	1.14	1.25	1.46
LASSO-ZERO	0.92	1.13	0.8	1.08	1.35	1.86	2.28	1.44
RIDGE	0.28	0.47	0.56	0.53	0.65	1.09	1.32	1.6
ELASTIC	0.26	0.34	0.36	0.59	0.72	1.51	1.04	1.52
ADAP-ELASTIC	0.27	0.35	0.36	0.57	0.7	1.49	1.03	1.53
SCAD	0.19	0.32	0.4	0.54	0.78	1.13	1.31	1.6
BMS	0.73	1.03	1.13	1.24	1.03	2.05	2.19	1.59
BMA	0.64	0.53	0.95	0.83	0.78	2.15	1.78	1.48
CSR	0.2	0.33	0.43	0.5	0.63	1.02	1.28	1.32
NEURAL-NET	0.25	0.39	0.39	0.42	0.55	0.86	**0.91**	1.09
RANDOM-FOREST	0.24	0.35	0.41	0.51	0.64	0.91	1.24	1.36
XG-BOOST	0.22	0.37	0.38	0.47	0.6	0.89	1.19	1.36

Table 8: Unconditional root-mean squared-error (8/5-digit core)
See Table 2 for model acronyms.

Once again, the results from the unconditional forecasting function differ to those that employ the conditional forecasting function, which are contained in appendix D.4, where it is suggested that the models that employ variable selection techniques provide results that may be superior to the benchmarks, while the nonlinear statistical learning models are also similar to what was reported for the four-digit data, but once again, they are slightly more impressive, where the neural network and boosting models provide more accurate forecasts for horizons that exceed three months. Similar results are provided in appendix E.4, which contains the MAPE statistics, while the Diebold and Mariano (1995) statistics that are contained in Table 9, suggest that there is no occasion where there is a significant difference in forecasting performance, in favour of the models that are competing with the random walk.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	**-1.99**	**-2.7**	**-2.98**	**-2.77**	**-2.09**	-1.36	-1.05	-0.94
AR-DIRECT	**-2.17**	**-2.05**	-1.93	**-2.03**	-1.41	-0.63	-0.36	-0.21
STOCH-VOL	**-2.65**	**-2.36**	**-2.28**	-1.77	-1.46	-0.62	-0.31	-0.18
BVAR	**-3.09**	**-2.78**	**-2.7**	**-2.31**	-1.95	-0.92	-0.44	-0.34
DIM	-1.78	**-2.6**	**-2.91**	**-3.02**	**-3.33**	**-2.21**	-1.25	-0.89
SARB	1.37	0.49	-0.68	-1.49	-1.82	-1.84	-0.84	-0.56
DFM-TF	-0.79	-0.16	-0.4	-0.16	-0.66	-0.86	**-2.38**	-1.32
DFM-3PRF	**-2.87**	**-1.98**	-1.89	-1.71	-1.85	**-2.26**	**-6.55**	-1.77
LASSO	**-3.32**	**-2.73**	-1.51	**-2.21**	**-5.8**	-1.6	-0.22	-0.73
LASSO-PSI	**-4.55**	**-3.1**	**-2.46**	**-2.96**	-1.11	-0.75	**-5.89**	-0.47
ADAP-LASSO	**-3**	**-2.87**	-1.4	**-2.06**	**-5.64**	-1.51	-0.19	-0.71
POST-LASSO	**-2.36**	-1.76	**-1.98**	-1.68	-1.81	**-6.37**	-1.6	-1.21
LASSO-ZERO	**-9.43**	**-3.49**	**-3.03**	**-2.89**	**-2.82**	**-3.89**	**-2.46**	**-2.18**
RIDGE	**-3.66**	**-4.11**	**-3.75**	**-2.06**	-1.94	-0.94	-0.52	-0.74
ELASTIC	**-3.15**	**-2.36**	-1.11	**-2.13**	**-3.7**	-1.27	-0.21	-0.69
ADAP-ELASTIC	**-3.22**	**-2.33**	-1.17	**-2.14**	**-3.66**	-1.18	-0.19	-0.68
SCAD	-1.62	**-2.3**	**-2.12**	**-2.24**	**-2.92**	**-2.08**	-1.3	-0.92
BMS	**-7.78**	**-4.97**	**-3.82**	**-3.86**	-1.93	**-2**	-1.13	-1.23
BMA	**-3.44**	**-2.75**	**-3.14**	**-3.29**	**-2.11**	**-3.16**	-0.86	-1.01
CSR	**-2.38**	**-2.12**	**-2.39**	**-2.09**	-1.74	**-4.76**	-1.07	-0.51
NEURAL-NET	**-2.64**	**-2.13**	-1.21	-0.91	-0.98	-0.37	0.06	-0.1
RANDOM-FOREST	**-2.58**	**-2.28**	**-2.42**	**-2.9**	**-2.88**	-0.95	-0.59	-0.63
XG-BOOST	**-2.67**	**-2.12**	-1.51	-1.56	**-3.34**	-0.59	-0.68	-0.66

Table 9: Diebold-Mariano statistics (8/5-digit core)

Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

4.5 Understanding the drivers of future inflationary pressures

To provide policymakers with an indication of sources of future inflationary pressure, we calculate Shapley values from the above statistical learning models. Figure 2 displays the results for the five largest and smallest Shapley values, which are provided by the XG boost model over a horizon of between one and six months ahead over the recursive out-of-sample period that extends over 48 months up until the end of March 2021. Note that in this case, the results suggest that the most significant upward pressure that is going to be placed on inflation originates from sunflower oil. Over the following three months, the popular press started to observe that sunflower oil was subject to significant price increases (c.f. Ledwaba (2021a) and Ledwaba (2021b)) and the Reserve Bank noted these drivers of inflationary pressure in its Monetary Policy Review (see South African Reserve Bank (2021)). This suggests that these models are able to correctly identify some of the future drivers of inflationary pressure.

Figure 2: Shapley values - headline inflation

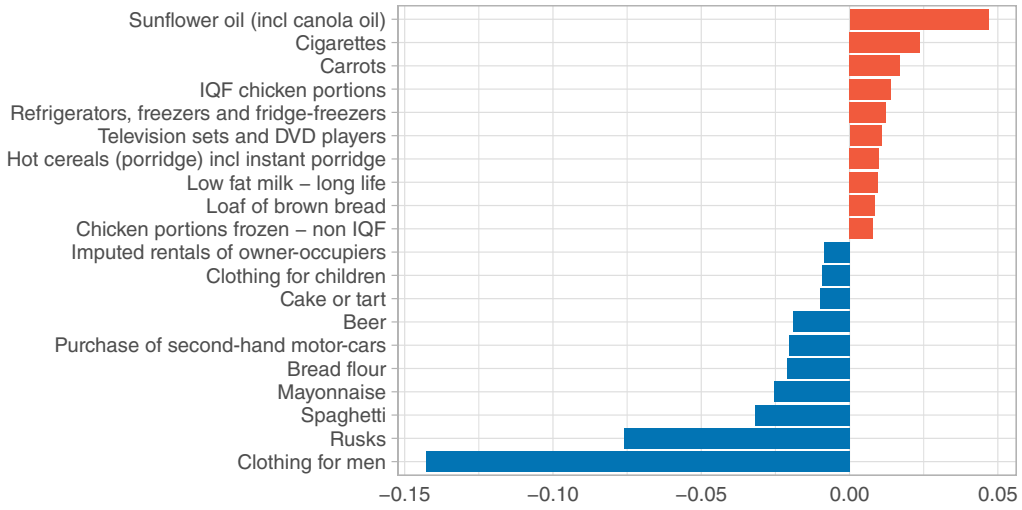
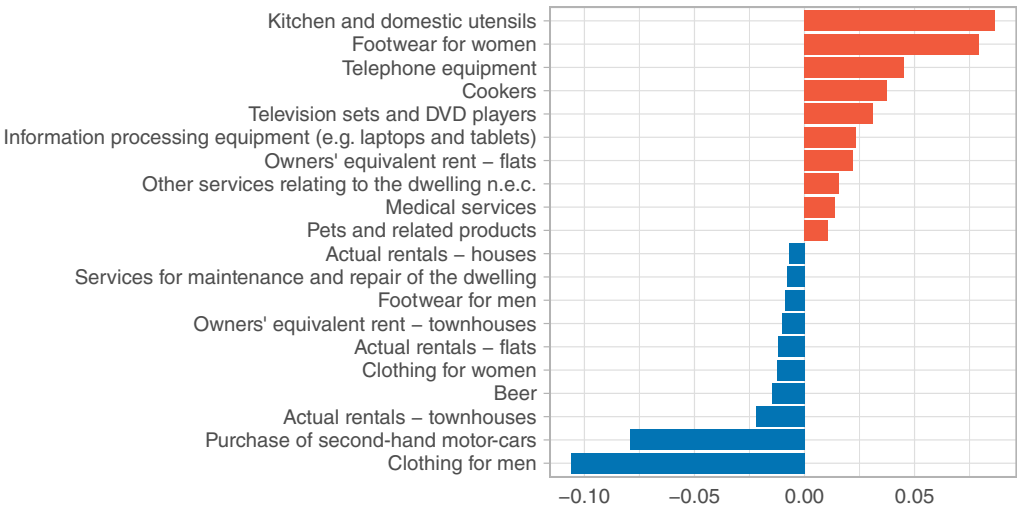


Figure 3: Shapley values - core inflation



Similar findings may be produced for core inflation, where the price of kitchen utensils increased, possibly as a result of the economic lockdown measures that would have forced more people to cook for themselves, thus stimulating demand for such items. Note that the use of more disaggregated price data in such a study would facilitate additional insight into the potential drivers of future inflationary pressure, which is an important consideration for policymakers. In addition, this type of analysis would also allow for policymakers to assess whether the source of future inflationary pressure is likely to be permanent or transitory. For example, after identifying that the sunflower oil is going to possibly contribute towards an increase in future inflation, policymakers would then be able to investigate whether this may be due to a temporary market shock, in which case they would not want to respond to it, or whether it is providing an early indicator of a permanent increase in inflationary pressure across a broad range of goods, which would necessitate a policy response.

4.6 Change in the inflationary trend

During periods of economic crisis, we would usually observe a decline in economic activity, which is usually followed by a decline in the demand for goods and services, and a subsequent reduction in consumer inflation. Following the onset of the Global Financial Crisis, such behaviour was present in most countries and it allowed for several central banks to make use of quantitative easing policies that sought to stimulate the subsequent economic recovery. Over this period of time, we experienced an initial downward trend in consumer inflation that persisted for an extended period of time, before the economic recovery eventually stimulated consumer demand to place upward pressure on prices.

Since the nonlinear statistical learning models could potentially use the information from a change in the inflationary trend to adjust the forecasts that they produce, while the traditional benchmark models would not incorporate this feature, we now investigate the degree to which these models may account for the change in trend inflation, which arose following the onset of the COVID pandemic. Previously, Stock and Watson (2010) have suggested that to account for the change in the inflationary trend one could augment the previous specification that they suggested in Stock and Watson (2007), with a stochastic trend that reacts to the unemployment recession gap, however, as is the case with most low- and middle-income countries, South Africa does not have a reliable measure for the unemployment recession gap, so as an alternative we now consider the forecasts that are generated by models that may potentially incorporate this nonlinear feature.

Figure 4 contains the results of forecasts that were generated for headline inflation, two years prior to the end of the sample (in March 2021). These forecasts pertain to horizons for over one and 24 steps ahead, when using either a conditional or unconditional forecasting function, which are displayed on the left and right panels, respectively. Note that when we have some information about the future evolution of inflation, then the neural network model, which has been used in this case is able to detect a change in the trend, while the random-walk remains constant and the autoregressive model converges on its mean, which is above 5%. However, when using the unconditional forecast function, where we do not have any information about future inflation then the neural network model does not appear to be able to correctly identify a decline in the trend. Figure 5 displays similar results for core inflation, when using either a conditional or unconditional forecasting function.

Figure 4: Forecasts from benchmark and statistical learning models - headline inflation

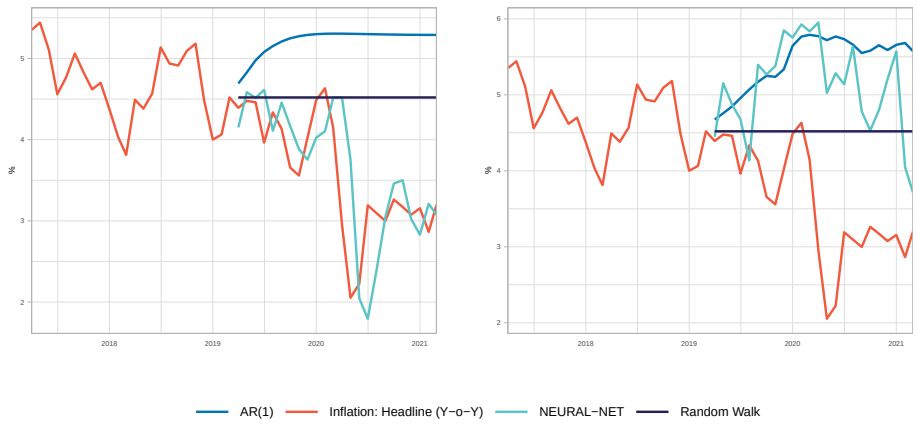
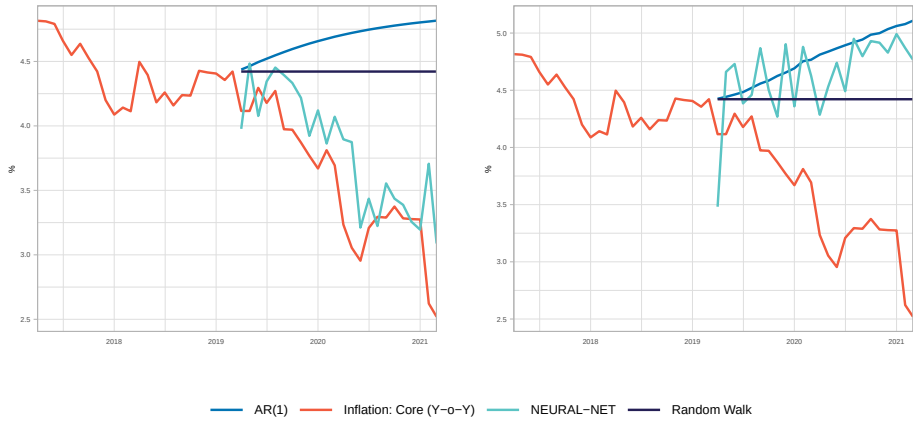


Figure 5: Forecasts from benchmark and statistical learning models - core inflation



5 Conclusion

We assess the potential predictive power of a number of different inflation forecasting models that may be applied to large datasets. We find that the models that employ variable selection techniques and nonlinear statistical learning techniques provide impressive results, when we make use of a conditional forecasting function that incorporates information about the future value of inflation. With the aid of this forecasting function, we also note that the use of models that seek to exploit any potential sparsity in the set of predictors provide relatively impressive results, when compared to the models that seek to summarise all of the available information. Over horizons that are longer than three months, the statistical learning models would also appear to provide results that are even more accurate than the sparse models, where the neural network and boosting models provide the most accurate results. These findings hold for both headline and core inflation. However, we have noted that all of these results are dependent upon the use of the conditional forecasting function, since when we use an unconditional forecasting function then the results of simple forecasting models continue to produce results that are in many cases superior to those of the statistical learning models. This is of importance to policymakers, as they would usually not be in a position where they are able to access the information that is assumed to be available when applying the conditional forecasting function. Hence, one would conclude that from a practical perspective, the use of statistical learning models in this particular setting, may not provide forecasts that are consistently superior to what is provided by a simple random walk.

Furthermore, the results suggest that for headline inflation the official central bank forecast that is presented to the MPC, which incorporates various sources of "off-model" information is more accurate than any of the other models, over the first three months. Similarly, over a one-month horizon the central bank forecast for core inflation is more accurate than any of the other models. Hence, the use of judgement has systematically improved the SARB forecasts over a short-term horizon. Another important finding relates to the use of more disaggregated data, where the results from the eight/five-digit levels are generally more accurate than when we report on the use of the four-digit data, which suggests that the use of more disaggregated data provides more desirable results. In particular, those models that are able to distinguish between information that may or may not be of potential use, are able to provide more accurate forecasts when they are applied to more disaggregated data. As has been shown, we can also use the output from the models to generate Shapley values, which provide policymakers with information that pertains to the drivers of future inflationary pressure. In addition, when we consider the relative performance of the benchmark models, which include a number of mean-reverting specifications, for the period that includes the effects of economic lockdowns and significant instability, we note that the statistical learning models are able to detect the decrease in the trend of the respective measures of inflation reasonably quickly, when we employ the conditional forecasting function. Subsequent research into the use of alternative sources of big data, as well as the potential use of alternative statistical learning model specifications, may provide more

promising forecasting results when we make use of an unconditional forecasting function.

References

Agrawal, A, Gans, J and Goldfarb, A. 2019. *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago.

Alpanda, S, Kotzé, K and Woglom, G. 2010a. 'The role of the exchange rate in a new Keynesian DSGE model for the South African economy'. *South African Journal of Economics* 78: 170–191.

Alpanda, S, Kotzé, K and Woglom, G. 2010b. 'Should central banks of small open economies respond to exchange rate fluctuations? The case of South Africa', *ERSA working paper no. 174*, Economic Research Southern Africa.

Alpanda, S, Kotzé, K and Woglom, G. 2011. 'Forecasting performance of an estimated DSGE model for the South African economy'. *South African Journal of Economics* 79: 50–67.

Athey, S. 2017. 'Beyond prediction: Using big data for policy problems'. *Science* 355: 483–485.

Athey, S. 2018. *The Impact of Machine Learning on Economics*, University of Chicago Press, 507–547.

Athey, S and Imbens, G W. 2019. 'Machine learning methods that economists should know about'. *Annual Review of Economics* 11: 685–725.

Bai, J. 2003. 'Inferential theory for factor models of large dimensions'. *Econometrica* 71: 135–171.

Bai, J and Ng, S. 2008. 'Large dimensional factor analysis'. *Foundations and Trends in Econometrics* 3: 89–163.

Baker, S R, Bloom, N, Davis, S J and Terry, S J. 2020a. 'Covid-induced economic uncertainty', *Working Paper 26983*, National Bureau of Economic Research.

Baker, S R, Farrokhnia, R A, Meyer, S, Pagel, M, Yannelis, C and Pontiff, J. 2020b. 'How Does Household Spending Respond to an Epidemic? Consumption during the 2020 COVID-19 Pandemic'. *Review of Asset Pricing Studies* 10: 834–862.

Balcilar, M, Gupta, R and Kotzé, K. 2015. 'Forecasting macroeconomic data for an emerging market with a nonlinear DSGE model'. *Economic Modelling* 44: 215–228.

Balcilar, M, Gupta, R and Kotzé, K. 2017. 'Forecasting South African macroeconomic variables with a Markov-switching small open-economy dynamic stochastic general equilibrium model'. *Empirical Economics* 53: 117–135.

Baldacci, E, Buono, D, Kapetanios, G, Krische, S, Marcellino, M, Mazzi, G L and Papailias, F. 2016. 'Big data and macroeconomic nowcasting: From data access to modelling', European Union, Eurostat, Luxembourg.

Bańbura, M, Giannone, D and Reichlin, L. 2010. 'Large Bayesian vector auto regressions'. *Journal of Applied Econometrics* 25: 71–92.

Barbieri, M M and Berger, J O. 2004. 'Optimal predictive model selection'. *The Annals of Statistics* 32: 870–897.

Bayarri, M J, Berger, J O, Forte, A and García-Donato, G. 2012. 'Criteria for Bayesian model choice with application to variable selection'. *The Annals of Statistics* 40: 1550–1577.

Belloni, A and Chernozhukov, V. 2013. 'Least squares after model selection in high-dimensional sparse models'. *Bernoulli* 19: 521–547.

Belloni, A, Chernozhukov, V, Fernández-Val, I and Hansen, C. 2017. 'Program evaluation and causal inference with high-dimensional data'. *Econometrica* 85: 233–298.

Belloni, A, Chernozhukov, V and Hansen, C. 2013. 'Inference on treatment effects after selection among high-dimensional controls'. *The Review of Economic Studies* 81: 608–650.

Belloni, A, Chernozhukov, V and Hansen, C. 2014. 'High-dimensional methods and inference on structural and treatment effects'. *Journal of Economic Perspectives* 28: 29–50.

Belloni, A, Chernozhukov, V and Wang, L. 2011. 'Square-root LASSO: pivotal recovery of sparse signals via conic programming'. *Biometrika* 98: 791–806.

Blumenstock, J. 2020. 'Machine learning can help get COVID-19 aid to those who need it most'. *Nature* .

Boivin, J and Ng, S. 2005. 'Understanding and comparing factor-based forecasts'. *International Journal of Central Banking* 1.

Botha, B, de Jager, S, Ruch, F and Steinbach, R. 2017. 'The Quarterly Projection Model of the SARB', *Working Paper 1*, South African Reserve Bank.

Breiman, L. 1996. 'Bagging predictors'. *Machine Learning* 24: 123–140.

Breiman, L. 2001. 'Random forests'. *Machine Learning* 45: 5–32.

Buckman, S R, Shapiro, A H, Sudhof, M and Wilson, D J. 2020. 'News Sentiment in the Time of COVID-19', *FRBSF Economic Letter 2020-08*, Federal Reserve Bank of San Francisco.

Carvalho, V M, Hansen, S, Ortiz, A, Ramón García, J, Rodrigo, T, Rodriguez Mora, S and Ruiz, J. 2020. 'Tracking the COVID-19 crisis with high-resolution transaction data', *CEPR Discussion Papers 14642*, C.E.P.R. Discussion Papers.

Castillo, I, Schmidt-Hieber, J and van der Vaart, A. 2015. 'Bayesian linear regression with sparse priors'. *The Annals of Statistics* 43: 1986–2018.

Castle, J L, Doornik, J A and Hendry, D F. 2021. 'The value of robust statistical forecasts in the COVID-19 pandemic'. *National Institute Economic Review* 256: 19–43.

Cavallo, A. 2020. 'Inflation with COVID consumption baskets', *Working Paper 27352*, National Bureau of Economic Research.

Chakrabarti, R, Heise, S, Melcangi, D, Pinkovskiy, M and Topa, G. 2020a. 'Did state reopenings increase consumer spending?', *Liberty street economics*, Federal Reserve Bank of New York.

Chakrabarti, R, Heise, S, Melcangi, D, Pinkovskiy, M and Topa, G. 2020b. 'How did state reopenings affect small businesses?', *Liberty street economics*, Federal Reserve Bank of New York.

Chen, J and Chen, Z. 2008. 'Extended Bayesian information criteria for model selection with large model spaces'. *Biometrika* 95: 759–771.

Chetty, R, Friedman, J N, Hendren, N, Stepner, M and Team, T O I. 2020. 'The economic impacts of COVID-19: Evidence from a new public database built using private sector data', *NBER Working Papers 27431*, National Bureau of Economic Research, Inc.

Chu, B, Huynh, K, Jacho-Chavez, D and Kryvtsov, O. 2018. 'On the evolution of the United Kingdom price distributions'. *Annals of Applied Statistics* 12: 2618–2646.

Clark, T E and West, K D. 2007. 'Approximately normal tests for equal predictive accuracy in nested models'. *Journal of Econometrics* 138: 291–311.

Creamer, K, Farrel, G and Rankin, N. 2012. 'What price-level data can tell us about pricing conduct in South Africa'. *South African Journal of Economics* 80: 490–509.

Creamer, K and Rankin, N. 2008. 'Price setting in South Africa 2001 to 2007 stylised facts using consumer price micro data'. *Journal of Development Perspectives* 4: 93–118.

Diebold, F X and Mariano, R S. 1995. 'Predictive accuracy'. *Journal of Business and Economic Statistics* 13: 253–263.

Doan, T, Litterman, R B and Sims, C. 1984. 'Forecasting and conditional projection using realistic prior distributions'. *Econometric Reviews* 3: 1–100.

Doerr, S, Gambacorta, L and Serena, J M. 2021. 'Big data and machine learning in central banking', *BIS Working Papers 930*, Bank of International Settlements.

Du Plessis, S, Du Rand, G and Kotzé, K. 2015. 'Measuring core inflation in South Africa'. *South African Journal of Economics* 83: 527–548.

Duarte, C and Rua, A. 2007. 'Forecasting inflation through a bottom-up approach: How bottom is bottom?' *Economic Modelling* 24: 941–953.

Eickmeier, S and Ziegler, C. 2008. 'How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach'. *Journal of Forecasting* 27: 237–265.

Elliott, G, Gargano, A and Timmermann, A. 2013. 'Complete subset regressions'. *Journal of Econometrics* 177: 357–373.

Elliott, G, Gargano, A and Timmermann, A. 2015. 'Complete subset regressions with large-dimensional sets of predictors'. *Journal of Economic Dynamics and Control* 54: 86–110.

Fan, J and Li, R. 2001. 'Variable selection via nonconcave penalized likelihood and its oracle properties'. *Journal of the American Statistical Association* 96: 1348–1360.

Faust, J and Wright, J H. 2013. 'Forecasting Inflation', in *Handbook of Economic Forecasting*, edited by G Elliott, C Granger and A Timmermann, Elsevier, Handbook of Economic Forecasting, (2): 2–56.

Florescu, D, Karlberg, M, Reis, F, Rey Del Castillo, P, Skaliotis, M and Wirthmann, A. 2014. 'Will 'big data' transform official statistics?', European Union, Eurostat, Luxembourg.

Forni, M, Hallin, M, Lippi, M and Reichlin, L. 2000. 'The generalized dynamic-factor model: Identification and estimation'. *The Review of Economics and Statistics* 82: 540–554.

Friedman, J, Hastie, T and Tibshirani, R. 2000. 'Additive logistic regression: a statistical view of boosting'. *The Annals of Statistics* 28: 337–374.

Friedman, J H. 2001. 'Greedy function approximation: A gradient boosting machine'. *The Annals of Statistics* 29: 1189–1232.

Fuhrer, J C. 2010. 'Inflation persistence', in *Handbook of Monetary Economics*, edited by B M Friedman and M Woodford, Elsevier, Handbook of Monetary Economics, 3 (9): 423–486.

Galvao, A B. 2021. 'The COVID-19 pandemic and macroeconomic forecasting: An introduction to the spring 2021 special issue'. *National Institute Economic Review* 256: 16–18.

George, E I and McCulloch, R E. 1993. 'Variable selection via Gibbs sampling'. *Journal of the American Statistical Association* 88: 881–889.

Giannone, D, Lenza, M and Primiceri, G E. 2021. 'Economic predictions with big data: the illusion of sparsity', *Working Paper Series 2542*, European Central Bank.

Goulet Coulombe, P, Marcellino, M and Stevanović, D. 2021. 'Can machine learning catch the COVID-19 recession?' *National Institute Economic Review* 256: 71–109.

Gupta, R and Kabundi, A. 2010. 'Forecasting macroeconomic variables in a small open economy: a comparison between small- and large-scale models'. *Journal of Forecasting* 29: 168–185.

Gupta, R and Kabundi, A. 2011. 'A large factor model for forecasting macroeconomic variables in South Africa'. *International Journal of Forecasting* 27: 1076–1088.

Gupta, R and Steinbach, R. 2013. 'A DSGE-VAR model for forecasting key South African macroeconomic variables'. *Economic Modelling* 33: 19–33.

Hahn, P R and Carvalho, C M. 2015. 'Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective'. *Journal of the American Statistical Association* 110: 435–448.

Hammer, C, Kostroch, D and Quiros-Romero, G. 2017. 'Big data: Potential, challenges and statistical implications', Washington, DC.

Hastie, T, Tibshirani, R and Wainwright, M. 2015. *Statistical Learning with Sparsity: The LASSO and Generalizations*, CRC Press, Boca Raton, FL.

Hubrich, K and Hendry, D F. 2005. 'Forecasting aggregates by disaggregates', *Computing in Economics and Finance 2005 270*, Society for Computational Economics.

Ibarra, R. 2012. 'Do disaggregated CPI data improve the accuracy of inflation forecasts?' *Economic Modelling* 29: 1305–1313.

Johnson, V E and Rossell, D. 2010. 'On the use of non-local prior densities in Bayesian hypothesis tests'. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72: 143–170.

Johnson, V E and Rossell, D. 2012. 'Bayesian model selection in high-dimensional settings'. *Journal of the American Statistical Association* 107: 649–660.

Joseph, A, Kalamara, E, Kapetanios, G and Potjagailo, G. 2020. 'Forecasting UK inflation bottom up', in *Nontraditional Data and Statistical Learning with Applications to Macroeconomics*, Bank of Italy and Federal Reserve Board.

Joseph, A, Kalamara, E, Kapetanios, G and Potjagailo, G. 2021. 'Forecasting UK inflation bottom up', *Staff Working Paper 915*, Bank of England.

Kanda, P T, Balcilar, M, Bahramian, P and Gupta, R. 2016. 'Forecasting South African inflation using non-linear models: a weighted loss-based evaluation'. *Applied Economics* 48: 2412–2427.

Kelly, B and Pruitt, S. 2013. 'Market expectations in the cross-section of present values'. *Journal of Finance* 68: 1721–1756.

Kelly, B and Pruitt, S. 2015. 'The three-pass regression filter: A new approach to forecasting using many predictors'. *Journal of Econometrics* 186: 294–316.

Klenow, P J and Malin, B A. 2010. 'Microeconomic evidence on price-setting', in *Handbook of Monetary Economics*, edited by B M Friedman and M Woodford, Elsevier, Handbook of Monetary Economics, 3 (6): 231–284.

Koop, G, McIntyre, S, Mitchell, J and Poon, A. 2021. 'Nowcasting 'true' monthly US GDP during the pandemic'. *National Institute Economic Review* 256: 44–70.

Ledwaba, K. 2021a. 'Cooking oil price soars as crop production slows', Sowetan Live. 21 June.

Ledwaba, K. 2021b. 'The rising price of cooking oil', Sowetan Live. 21 June.

Lee, J D, Sun, D L, Sun, Y and Taylor, J E. 2016. 'Exact post-selection inference with the LASSO'. *Annals of Statistics* 44: 907–927.

Liang, F, Paulo, R, Molina, G, Clyde, M A and Berger, J O. 2008. 'Mixtures of g priors for Bayesian variable selection'. *Journal of the American Statistical Association* 103: 410–423.

Litterman, R B. 1986a. 'Forecasting with Bayesian vector autoregressions: Five years of experience'. *Journal of Business & Economic Statistics* 4: 25–38.

Litterman, R B. 1986b. 'A statistical approach to economic forecasting'. *Journal of Business & Economic Statistics* 4: 1–4.

Liu, G D, Gupta, R and Schaling, E. 2009. 'A New-Keynesian DSGE model for forecasting the South African economy'. *Journal of Forecasting* 28: 387–404.

Lundberg, S and Lee, S I. 2017. 'A unified approach to interpreting model predictions', in *Advances in Neural Information Processing Systems*, volume 30, 4765 – 4774.

McCracken, M W. 2007. 'Asymptotics for out of sample tests of granger causality'. *Journal of Econometrics* 140: 719–752.

McCracken, M W and Ng, S. 2016. 'FRED-MD: A monthly database for macroeconomic research'. *Journal of Business & Economic Statistics* 34: 574–589.

McCullagh, P and Polson, N G. 2018. 'Statistical sparsity'. *Biometrika* 105: 797–814.

Medeiros, M C, Vasconcelos, G F R, Veiga, Á and Zilberman, E. 2021. 'Forecasting inflation in a data-rich environment: The benefits of machine learning methods'. *Journal of Business & Economic Statistics* 39: 98–119.

Mehrhoff, J. 2017. 'Central banks' use of and interest in big data', in *Big Data*, edited by Bank for International Settlements, Bank for International Settlements, IFC Bulletins.

Mitchell, T J and Beauchamp, J J. 1988. 'Bayesian variable selection in linear regression'. *Journal of the American Statistical Association* 83: 1023–1032.

Mullainathan, S and Spiess, J. 2017. 'Machine learning: An applied econometric approach'. *Journal of Economic Perspectives* 31: 87–106.

Nakamura, E and Steinsson, J. 2013. 'Price rigidity: Microeconomic evidence and macroeconomic implications'. *Annual Review of Economics* 5: 133–163.

OECD. 2020. 'Using artificial intelligence to help combat COVID-19', *Technical report*, Organisation for Economic Co-operation and Development.

Petrella, I, Santoro, E and Simonsen, L P. 2019. 'Time-varying price flexibility and inflation dynamics', *EMF Research Papers 28*, Economic Modelling and Forecasting Group.

Rossell, D. 2021. 'Concentration of posterior probabilities and normalized L_0 criteria in regression'. *Bayesian Analysis* 1: 1–27.

Rossell, D and Telesca, D. 2017. 'Nonlocal priors for high-dimensional estimation'. *Journal of the American Statistical Association* 112: 254–265.

Rousseau, J and Szabo, B. 2020. 'Asymptotic frequentist coverage properties of Bayesian credible sets for sieve priors'. *Annals of Statistics* 48: 2155–2179.

Ruch, F, Balcilar, M, Gupta, R and Modise, M P. 2020. 'Forecasting core inflation: the case of South Africa'. *Applied Economics* 52: 3004–3022.

Ruch, F, Rankin, N and du Plessis, S. 2016a. 'Decomposing inflation using micro-price data: Sticky-price inflation', *South African Reserve Bank Working Paper Series 7354*, South African Reserve Bank.

Ruch, F, Rankin, N and du Plessis, S. 2016b. 'Decomposing inflation using micro price level data: South Africa's pricing dynamics', *Working papers*, South African Reserve Bank.

Schmitt-Grohé, S and Uribe, M. 2004. 'Solving dynamic general equilibrium models using a second order approximation of the policy function'. *Journal of Economic Dynamics and Control* 28: 755–75.

Schumacher, C. 2007. 'Forecasting German GDP using alternative factor models based on large datasets'. *Journal of Forecasting* 26: 271–302.

Scott, J G and Berger, J O. 2010. 'Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem'. *The Annals of Statistics* 38: 2587–2619.

Shapiro, A H, Sudhof, M and Wilson, D J. 2017. 'Measuring news sentiment', *Working Paper 2017-01*, Federal Reserve Bank of San Francisco.

Shapley, L S. 1953. 'A value for n-person games', in *Contributions to the Theory of Games*, edited by H Kuhn and A Tucker, Princeton University Press, Princeton, Annals of Mathematical Studies, (20): 307–317.

Smal, D, Pretorius, C and Ehlers, N. 2007. 'The core forecasting model of the South African Reserve Bank', *Working Paper WP/07/02*, South African Reserve Bank.

South African Reserve Bank. 2021. 'Monetary Policy Review (April)', *Technical report*, South African Reserve Bank.

Statistics South Africa. 2007. 'Shopping for two: The CPI new basket parallel survey - Results and comparisons with published CPI data', *Technical report*, Statistics South Africa.

Statistics South Africa. 2017a. 'Consumer price index: The South African CPI sources and methods manual', *Technical report*, Statistics South Africa.

Statistics South Africa. 2017b. 'Introduction of new weights and basket for the consumer price index', *Technical report*, Statistics South Africa.

Steinbach, R, Mathuloe, P and Smit, B. 2009. 'An open economy New Keynesian DSGE model of the South African economy', *Working Papers 3431*, South African Reserve Bank.

Stock, J H and Watson, M W. 2002a. 'Forecasting using principal components from a large number of predictors'. *Journal of the American Statistical Association* 97: 1167–1179.

Stock, J H and Watson, M W. 2002b. 'Macroeconomic forecasting using diffusion indexes'. *Journal of Business & Economic Statistics* 20: 147–162.

Stock, J H and Watson, M W. 2006. 'Chapter 10 forecasting with many predictors', Elsevier, Handbook of Economic Forecasting, 1: 515–554.

Stock, J H and Watson, M W. 2007. 'Why has U.S. Inilation become harder to forecast?' *Journal of Money, Credit and Banking* 39: 3–33.

Stock, J H and Watson, M W. 2010. 'Modeling inflation after the crisis', *Working Paper 16488*, National Bureau of Economic Research.

Stock, J H and Watson, M W. 2020. 'Slack and cyclically sensitive inflation'. *Journal of Money, Credit and Banking* 52: 393–428.

Taylor, J and Tibshirani, R. 2018. 'Post-selection inference for L_1-penalized likelihood models'. *Canadian Journal of Statistics* 46: 41–61.

Tibshirani, R. 1996. 'Regression shrinkage and selection via the LASSO'. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 58: 267–288.

Tissot, B. 2019. 'Big data for central banks', in *The use of big data analytics and artificial intelligence in central banking*, Bank for International Settlements.

United Nations Global Pulse (UNGP). 2012. 'Big data for development: Challenges and opportunities'.

Varian, H R. 2014. 'Big data: New tricks for econometrics'. *Journal of Economic Perspectives* 28: 3–28.

Wibisono, O, Ari, H D, Widjanarti, A, Zulen, A A and Tissot, B. 2019. *The use of big data analytics and artificial intelligence in central banking*, IFC Bulletins, Bank for International Settlements.

Woglom, G. 2005. 'Forecasting South African inflation'. *The South African Journal of Economics* 73: 302–320.

Wolters, M H and Tillmann, P. 2015. 'The changing dynamics of US inflation persistence: a quantile regression approach'. *Studies in Nonlinear Dynamics & Econometrics* 19: 161–182.

World Bank. 2014. 'Central America: Big data in action for development', Washington, DC.

Zellner, A. 1986. 'On assessing prior distributions and Bayesian regression analysis with g-prior distributions', in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, edited by P Goel and A Zellner, Elsevier, New York, 233–243.

Zou, H. 2006. 'The adaptive LASSO and its oracle properties'. *Journal of the American Statistical Association* 101: 1418–1429.

A Benchmark models

A.1 Autoregressive and random walk models

The traditional first-order autoregressive model takes the form, $y_{i+h} = \mu_0 + \phi_1 y_{i-1} + \varepsilon_i$, where the future value of a variable is related to previous observations and random disturbances, which is denoted by $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$. Hence, the forecast function for this model would be expressed as, $E_i[y_{i+h}|y_i] = y_i\hat{\phi}_1 + \hat{\mu}_0$, where the coefficient estimates are derived from data that is available at time i. The random walk model assumes that the last observed value of the process should be used for all the future predicted values. Therefore, it has a forecast function that could be expressed as, $E_i[y_{i+h}|y_i] = y_i$. In addition to making use of the traditional forecast function for the autoregressive model, we also make use of the direct forecast function, which is described in more detail below.

A.2 Stochastic volatility model

Stock and Watson (2007) suggest that the use of an unobserved components model with stochastic volatility provide relatively accurate forecasts of inflation. We use a similar specification for this model, which also incorporates an additional stochastic term that is used to explain the unexpected shocks to the volatility process. The parameters in the model and the unobserved components take the form of random variables and are estimated with Bayesian methods, in the model that is specified as:

$$\begin{aligned} y_i &= \mu_0 + \phi_1 y_{i-1} + a_i, && (33)\\ a_i &= \sigma_i \varepsilon_i && (34)\\ \log(\sigma_i^2) &= \alpha_0 + \alpha_1 \log(\sigma_{i-1}^2) + \upsilon_i. && (35) \end{aligned}$$

In this case, ε_i and υ_i are separate $i.i.d.\ N(0,1)$ processes. Beta priors are then used for ϕ_1 and α_1, while μ_t and α_0 take on uniform priors.

A.3 Large BVAR model

Litterman (1986a,b) and Doan et al. (1984), describe the use of what has become a popular macroeconomic forecasting model that applies Bayesian methods to impose greater degrees of shrinkage on all but the first lag of the target variable. Bańbura et al. (2010) extend this framework, where the degree of shrinkage in model is also a function of the cross-sectional dimension, to provide impressive results in a setting where the number of predictors is relatively large. This model would make use of the traditional vector autoregressive structure:

$$\boldsymbol{z}_i = \boldsymbol{\mu} + \beta_1 \boldsymbol{z}_{i-1} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim i.i.d.N(0, \sigma^2), \tag{36}$$

where $\boldsymbol{z}$ contains both the target variable and the predictors, such that $\boldsymbol{z} = \{\boldsymbol{y}, \boldsymbol{X}\}'$. Note that when applying the methodology of Bańbura et al. (2010), as p increases, the priors that are assigned to variables that do not relate lags of the target variable are more closely centred around zero, which implies that the forecasting performance of the model should be similar to those that are provided by an AR(1) when the number of predictors is extremely large. The motivation for employing this methodology is to reduce the perils of over-fitting and for this specification of the model, the forecast function would take the form $E_i[\boldsymbol{z}_{i+h}|\boldsymbol{z}_i] = \boldsymbol{z}_i'\hat{\beta}_1 + \hat{\boldsymbol{\mu}}$.

A.4 Disaggregated inflation model

The near-term inflation forecast by the SARB is largely dependent upon the predictions of the DIM, which makes use of monthly data that is provided to the public by StatsSA on the 46 components that are contained in the consumption basket. These forecasts are derived from a seasonal averaging technique for the percentage change in the index that is derived for each of the individual components in the basket, which is then used to generate a

projection for each of these components. To provide a single forecast for the index, which may relate to headline or core inflation, the reported weights for the consumption basket items are applied to the levels of each individual forecast for the CPI components. Therefore, to summarise this model, we make use of the notation j, which represents the individual components of the CPI, while $x_{i,j}$ is the percentage change of the CPI components at time i. Thereafter, let f be the frequency of observations, while n refers to total number of observations in $x_{i,j}$. A simple seasonal method that may be used to forecast the components of CPI may then take the following form:

$$E_i\left[y_{i+h}|x_{i,j}\right] = \frac{1}{p}\sum_{j=\lceil\frac{mod(h,f)}{f}\rceil}^{p} x_{i+mod(h,f)-fj,j}, \text{ where } n = \lfloor\frac{i+mod(h,f)-1}{f}\rfloor, \tag{37}$$

where $h = \{1,2,3,\ldots\}$. Hence, the projected percentage change in inflation is equal to the average of historical percentage change for a particular seasonal period, where by way of example, the forecast for inflation in April is equal to the weighted average of the historical percentage change that arose from March to April for each component of CPI.

Prior to submitting the results from the DIM forecast to the MPC, the forecasts for a number of CPI components are supplemented with "off-model" information. By way of example, information relating to the electricity tariff agreements between the National Energy Regulator of South Africa (NERSA) and the national power utility is used to adjust the projections that relate to the future price of energy, while forecasts for fuel prices are based on assumptions that relate to the predicted path of oil prices and the exchange rate, which are consistent with various high-frequency data updates and the forecasts of the quarterly projection model (QPM).[18] Note also that as the MPC meetings take place around the twentieth of each month, the central bank economists are able to make use of within-month observed values for certain predictors. For example, before submitting the forecast for the month of January, economists could make use of the regulated fuel prices that have been observed between the first and twentieth of January to amend their inflation forecast. Hence, when comparing the short-term forecasts of the SARB to those of other models, we should be conscious of this advantage, which is incorporated along with the additional off-model information that is utilised by the central bank economists. The degree to which this "judgement" influences the forecasts that are submitted to the MPC is quantified after we compare the forecasts of the DIM with the official forecasts that are submitted to the MPC.

[18] The QPM is the forecasting model of the SARB used to inform monetary policy, and is responsible for generating quarterly predictions. For further details on the features of this model, see Botha et al. (2017).

B Dimensionality reduction models

Principal component analysis (PCA) seeks to summarise all the available information that is contained in the predictors, after considering the degree of covariation in the combined set of variables. Therefore, the principal components are orthogonal variables that represent linear combinations of the original predictors, where the greatest amount of information is contained with the first principle component and the remaining components contain decreasing amounts of information. When applied in a setting that makes use of time series variables that may be subject to a certain degree of serial correlation, the dynamic factor model (DFM) is usually employed, which follows the seminal work of Forni et al. (2000), Stock and Watson (2002a,b) and Bai (2003). A general characterisation of the model would take the form:

$$y_{i+h} = \beta_0 + \beta' \boldsymbol{F}_i + \eta_{i+1}, \tag{38}$$

$$\boldsymbol{x}_i = \Phi \boldsymbol{F}_i + \boldsymbol{\varepsilon}_i, \tag{39}$$

where F_i represents the common driving forces in the regressors, β' is the vector of loadings, measuring the effects of the factors on inflation, while x_i is a set of p weakly stationary variables, that are also driven by F_t via the loading matrix, Φ. Stock and Watson (2002a,b) describe the conditions for the factors, factor loadings, residuals, permissible amount of temporal dependence and the degree of cross-sectional dependence. Reviews of the more recent literature on applications that make use of DFMs are contained in Boivin and Ng (2005), Stock and Watson (2006), Eickmeier and Ziegler (2008) and Schumacher (2007).

While there is extensive literature on the use of different methods of estimation for these models, which may be applied to large datasets, we confine ourselves to the use of PCA based estimators, due to their computational simplicity and their impressive empirical performance in previous studies. In particular, we make use of two different methodologies, which include the three pass regression filter of Kelly and Pruitt (2013, 2015) and an amended variant of the target factor approach, which is described in Bai and Ng (2008). Since these techniques focus on variable combination rather than variable selection, they have been termed "*dense*" models.

To implement the target factors approach, which seeks to make use of the principal components for the variables that have high prediction power, we initially regress y_i on x_{i-h}, where lagged values of y_i act as controls and are included in x_{i-h}. Before estimating the factors, we remove those predictors that are uncorrelated with the target variable, where the total number of variables that are included in the set of potential predictors is limited to be slightly less than the number of available observations. This ensures that we have sufficient degrees of freedom to provide estimates for the principal components.

C Estimation results

Table 10 contains selected in-sample summary statistics for the different models, where we have used the full sample of available four-digit data for headline inflation. Note that the model that imposes an L_0 penalty, which is labelled LASSO-ZERO, has a similar coefficient of determination to the linear regression model, however, it only makes use of 12 explanatory variables.

Model	Coefficient incl	Coefficient excl	MSE	R^2
LASSO	44	3	0.008	0.995
LASSO-PSI	45	2	0.008	0.995
ADAP-LASSO	46	1	0.008	0.995
POST-LASSO	16	31	0.026	0.982
LASSO-ZERO	12	35	0.016	0.99
RIDGE	47	0	0.008	0.995
ELASTIC	45	2	0.008	0.995
ADAP-ELASTIC	45	2	0.008	0.995
SCAD	42	5	0.008	0.995
BMS	18	29	0.012	0.992
BMA	42	5	0.008	0.995

Table 10: In-sample summary statistics: four-digit headline inflation

Table 11 contains selected in-sample summary statistics for the different models, where we have used the full sample of available 8/5-digit data for headline inflation. Note that in most cases, a large number of potential regressors are excluded, where the model that imposes the L_0 penalty makes use of only eleven of the two-hundred and ten regressors to provide a near perfect fit.

Model	Coefficient incl	Coefficient excl	MSE	R^2
LASSO	83	128	0.003	0.998
LASSO-PSI	82	129	0.002	0.999
ADAP-LASSO	76	135	0.003	0.998
POST-LASSO	30	181	0.013	0.989
LASSO-ZERO	10	201	0.014	0.988
RIDGE	207	4	0	1
ELASTIC	106	105	0.001	0.999
ADAP-ELASTIC	97	114	0.001	0.999
SCAD	16	195	0.036	0.969
BMS	13	198	0.012	0.99
BMA	54	157	0.004	0.997

Table 11: In-sample summary statistics: eight/five-digit headline inflation

Table 12 contains selected in-sample summary statistics for the different models, where we have used the full sample of available four-digit data for core inflation. Note that the Bayesian model that makes use of a non-local prior has a similar coefficient of determination to the linear regression model, however, it only makes use of 24 explanatory variables.

Model	Coefficient incl	Coefficient excl	MSE	R^2
LASSO	33	1	0.006	0.996
LASSO-PSI	31	3	0.006	0.996
ADAP-LASSO	31	3	0.006	0.995
POST-LASSO	13	21	0.026	0.981
LASSO-ZERO	17	17	0.008	0.994
RIDGE	34	0	0.006	0.996
ELASTIC	33	1	0.006	0.996
ADAP-ELASTIC	33	1	0.006	0.996
SCAD	29	5	0.006	0.996
BMS	23	11	0.007	0.995
BMA	34	0	0.006	0.996

Table 12: In-sample summary statistics: four-digit core inflation

Table 13 contains selected in-sample summary statistics for the different models, where we have used the full sample of available 8/5-digit data for core inflation. Note that in most cases, a large number of potential regressors are excluded, where the model that imposes the L_0 penalty makes use of only sixteen of the eighty-three regressors to provide a near perfect fit.

Model	Coefficient incl	Coefficient excl	MSE	R^2
LASSO	60	30	0.004	0.995
LASSO-PSI	73	17	0.002	0.997
ADAP-LASSO	55	35	0.004	0.995
POST-LASSO	16	74	0.022	0.972
LASSO-ZERO	8	82	0.019	0.976
RIDGE	88	2	0.002	0.997
ELASTIC	74	16	0.003	0.996
ADAP-ELASTIC	66	24	0.003	0.996
SCAD	72	18	0.003	0.997
BMS	22	68	0.007	0.991
BMA	65	25	0.003	0.997

Table 13: In-sample summary statistics: eight/five-digit core inflation

D Conditional forecasting results

Following Medeiros et al. (2021), we also make use of their conditional forecasting function, which utilises future information about the value of inflation that will be realised in the month prior to what we are looking to forecast. To explain how this forecasting function is applied in practice consider the example where we are looking to generate a 12-step-ahead forecast for the month of January 2021. This approach would make use of the observed value of inflation in December 2020, which would be regressed on values for all the potential predictors that were observed in December 2019. Thereafter, we would use these coefficients and multiply them by the values for all the potential predictors that were observed in January 2020 to obtain a forecast for January 2021. This approach differs to the case where we make use of the unconditional forecast function, which assumes that we need to generate a 12-step-ahead forecast for the month of January 2021, where we only have information for any of the variables for January 2020 (i.e. we do not have any information about what the rate of inflation will be in December 2020).

D.1 Four-digit data: headline inflation

Table 14 contains the out-of-sample RMSE for the different models.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.41	0.68	0.82	0.93	1.09	1.68	1.75	1.68
AR-DIRECT	0.42	0.7	0.85	0.97	1.12	1.55	1.66	1.69
STOCH-VOL	0.42	0.64	0.76	0.84	0.91	1.21	1.31	1.38
RAND-WALK	0.41	0.64	0.76	0.84	0.9	1.19	1.33	1.45
BVAR	0.6	0.79	0.93	1.04	1.2	1.41	1.45	1.47
DIM	0.36	0.59	0.72	0.82	0.97	1.49	1.61	1.69
SARB	**0.15**	**0.25**	**0.37**	0.57	0.84	1.42	1.59	1.63
LINEAR	0.5	0.61	0.69	0.61	0.74	0.65	0.75	0.79
DFM-TF	0.43	0.68	0.83	0.95	1.1	1.58	1.55	1.4
DFM-3PRF	0.45	0.66	0.76	0.81	0.98	1.21	1.32	1.47
LASSO	0.4	0.52	0.55	0.57	0.68	0.59	0.63	0.55
LASSO-PSI	0.6	1.42	1.19	1.08	1.51	1.7	2.21	2.84
ADAP-LASSO	0.4	0.53	0.57	0.57	0.7	0.59	0.63	0.54
POST-LASSO	0.44	0.61	0.58	0.69	0.8	0.96	0.86	0.99
LASSO-ZERO	0.42	0.62	0.67	0.65	0.72	0.69	0.61	0.7
RIDGE	0.47	0.61	0.59	0.59	0.67	0.59	0.68	0.56
ELASTIC	0.4	0.54	0.57	0.57	0.68	0.58	0.64	**0.52**
ADAP-ELASTIC	0.37	0.55	0.59	0.59	0.69	0.6	0.63	0.52
SCAD	0.44	0.64	0.62	0.62	0.66	0.77	0.64	0.84
BMS	0.44	0.62	0.66	0.59	0.69	0.69	0.62	0.72
BMA	0.43	0.66	0.64	0.63	0.67	0.65	0.58	0.67
CSR	0.41	0.61	0.68	0.71	0.74	1.06	1.17	1.14
NEURAL-NET	0.51	0.52	0.52	**0.5**	**0.54**	**0.58**	0.59	0.64
RANDOM-FOREST	0.51	0.59	0.65	0.69	0.72	0.67	0.65	0.75
XG-BOOST	0.43	0.5	0.51	0.54	0.59	0.59	**0.49**	0.66

Table 14: Conditional root-mean squared-error (four-digit headline)
See Table 2 for model acronyms.

When we compare the results of the "dense" models relative to the "sparse" models we note that the models that make use of variable selection techniques are responsible for more accurate forecasts, over all horizons. In addition, while the DFMs do not appear to provide any advantage over the benchmarks, the sparse models do provide results that are at least equivalent, but in most cases superior to those of the autoregressive, stochastic volatility and random walk benchmarks. These models also provide results that are superior to the DIM at all

but the 1-step-ahead horizon, where the RMSE is about half that of the benchmarks over a horizon that is 12 months or more. The results for the nonlinear statistical learning models are in most cases similar to those of the sparse models, however, over longer horizons the neural network and boosting models provide forecasts that are slightly more accurate. In this case, the neural network provides the lowest RMSE at the 4-, 6-, 12- and 24-step ahead-horizon.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	-0.2	-1.65	-1.21	-1.26	-1.41	-0.79	-0.5	-0.38
AR-DIRECT	0.06	-1.36	-1.94	**-2.2**	-1.18	-0.58	-0.38	-0.25
STOCH-VOL	-0.19	-0.03	0.07	0.06	-0.1	-0.07	0.04	0.17
BVAR	**-2.21**	-1.77	-1.94	-1.6	-1.16	-0.46	-0.19	-0.03
DIM	-1.67	-1.16	-1.01	-1.12	-1.3	-0.79	-0.5	-0.44
SARB	*3.67*	*2.3*	1.84	1.88	0.35	-0.61	-0.56	-0.37
LINEAR	-1.46	0.46	0.54	1.73	1.47	*2.29*	*4.23*	*4.21*
DFM-TF	-0.77	-0.63	-0.86	-0.93	-0.8	-0.83	-0.31	0.11
DFM-3PRF	-1.36	-0.5	0.08	0.29	-0.43	-0.04	0.01	-0.12
LASSO	0.27	1.1	1.49	*2.11*	1.68	*2.3*	*3.69*	*6.11*
LASSO-PSI	**-3.29**	**-2.2**	**-2.43**	-1.84	-1.6	**-2.75**	-1.28	-1.22
ADAP-LASSO	0.38	1.05	1.34	1.91	1.43	*2.32*	*3.65*	*6.1*
POST-LASSO	-0.97	1.25	1.54	1.83	0.68	0.76	*4.04*	*3.73*
LASSO-ZERO	-0.32	0.55	0.87	1.8	1.09	*2.38*	*5.45*	*5.48*
RIDGE	-0.84	0.29	1.07	*2.01*	1.33	*2.39*	*2.7*	*5.93*
ELASTIC	0.41	0.91	1.36	*2.15*	1.52	*2.44*	*3.63*	*6.19*
ADAP-ELASTIC	0.88	0.78	1.2	*1.98*	1.41	*2.54*	*4.19*	*6.23*
SCAD	-1.1	-0.01	1.51	1.63	1.76	1.79	*5.3*	*5.17*
BMS	-1.51	0.39	0.93	1.89	1.28	*2.29*	*4.96*	*5.45*
BMA	-0.42	-0.31	0.85	1.43	1.64	*2.22*	*5.12*	*5.19*
CSR	0.43	1.03	1.86	*2.19*	1.39	0.41	0.35	*2.54*
NEURAL-NET	**-2.17**	1.5	1.52	1.74	*2.38*	*2.3*	*4.87*	*5.62*
RANDOM-FOREST	-1.26	1.3	1.51	*2.13*	1.1	1.88	*3.62*	*5.38*
XG-BOOST	-0.72	1.8	1.83	*2.15*	*1.97*	*2.1*	*5.81*	*5.72*

Table 15: Conditional Diebold-Mariano statistics (four-digit headline)
Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

Turning our attention to the Diebold-Mariano statistics, which are contained in Table 15 we can confirm that the difference in forecasting performance, relative to the random walk model, are in most cases significantly different from zero, in favour of the competing models. This is indicated by the relatively large positive values, that are highlighted in italics.

D.2 Eight/five-digit data: headline inflation

Table 16 contains the out-of-sample RMSE for the different models, where most of the results that pertain to the benchmarks are the same as what was provided for the four-digit data. Similar to what was provided previously, the results for all of the models that employ variable selection techniques are superior to the benchmark models that do not use "off-model" information when the horizon is greater than three months. In addition, note that the RMSEs for these models are slightly lower when using the more disaggregated data, which would suggest that the combined use of more disaggregated data and variable selection techniques allows for an improved forecasting performance, as it would discard some of the noise that may be included in the variables when they are subject to greater degrees of aggregation.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.41	0.79	0.93	1.01	1.23	1.63	1.68	1.62
AR-DIRECT	0.41	0.69	0.83	0.94	1.08	1.41	1.47	1.49
STOCH-VOL	0.42	0.65	0.76	0.84	0.92	1.23	1.34	1.41
RAND-WALK	0.41	0.64	0.76	0.84	0.9	1.19	1.33	1.45
BVAR	0.51	0.68	0.79	0.9	1.09	1.45	1.61	1.72
DIM	0.36	0.59	0.72	0.82	0.97	1.49	1.61	1.69
SARB	**0.15**	**0.25**	**0.37**	0.57	0.84	1.42	1.59	1.63
DFM-TF	0.43	0.73	0.87	1.03	1.38	1.69	1.73	1.43
DFM-3PRF	0.46	0.68	0.78	0.92	1.17	1.43	1.71	2.05
LASSO	0.43	0.46	0.44	0.49	0.51	0.52	0.58	0.56
LASSO-PSI	1.43	1.21	1.79	1.92	1.85	2.54	2.07	2.12
ADAP-LASSO	0.43	0.46	0.44	0.51	0.52	0.48	0.58	0.54
POST-LASSO	0.45	0.64	0.62	0.7	0.79	0.76	0.77	0.82
LASSO-ZERO	0.53	1.51	2.18	2.38	4.43	1.49	2.89	2.78
RIDGE	0.37	0.39	0.43	**0.43**	**0.46**	0.58	0.5	0.6
ELASTIC	0.42	0.44	0.41	0.47	0.5	0.49	0.51	0.55
ADAP-ELASTIC	0.41	0.46	0.42	0.48	0.51	0.48	0.54	0.55
SCAD	0.4	0.55	0.63	0.63	0.67	0.66	0.65	0.73
BMS	1.76	2.48	1.78	2.7	1.63	2.14	1.76	2.62
BMA	2.13	2.03	1.94	3.11	1.75	2.97	1.62	3.25
CSR	0.41	0.6	0.69	0.76	0.81	0.85	0.95	1.02
NEURAL-NET	0.49	0.48	0.5	0.61	0.53	**0.47**	**0.49**	**0.5**
RANDOM-FOREST	0.58	0.66	0.72	0.77	0.75	0.68	0.79	0.8
XG-BOOST	0.46	0.55	0.6	0.68	0.67	0.67	0.73	0.69

Table 16: Root-mean squared-error: 8/5-digit headline
See Table 2 for model acronyms.

And then lastly, the results for the nonlinear statistical learning models are once again superior to the models that employ variable selection techniques, when the horizon is greater than three months ahead. Note also that the results for the neural network are more impressive when using 8/5-digit data (than when using four-digit data), and they have reduced to less than a third of the SARB forecast over a 24-month horizon.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.11	-1.38	**-2.03**	**-2.47**	-1.36	-0.67	-0.49	-0.26
AR-DIRECT	0.02	-1.06	-0.86	-0.9	-0.9	-0.46	-0.21	-0.06
STOCH-VOL	-0.37	-0.1	-0.03	-0.13	-0.27	-0.15	-0.03	0.1
BVAR	-1.42	-0.53	-0.54	-0.69	-1.04	-0.58	-0.39	-0.22
DIM	*2.19*	1.46	0.91	0.46	-0.61	-0.78	-0.5	-0.43
SARB	*3.67*	*2.3*	1.84	1.88	0.35	-0.61	-0.56	-0.37
DFM-TF	-0.69	-0.88	-1.3	**-2.16**	-1.93	-1.66	-0.49	0.06
DFM-3PRF	-1.32	-0.83	-0.17	-0.89	-1.88	-0.44	-1.05	-1.16
LASSO	-0.36	1.96	*2.18*	*2.4*	*2.86*	*2.52*	*5.73*	*6.22*
LASSO-PSI	**-3.37**	**-3.36**	**-4.22**	**-2.7**	-1.85	**-3.11**	**-4.58**	-0.93
ADAP-LASSO	-0.28	*1.96*	*2.16*	*2.37*	*2.81*	*2.66*	*5.68*	*6.19*
POST-LASSO	-1.25	0.03	1.7	*2.81*	0.54	1.56	*3.71*	*5.19*
LASSO-ZERO	**-2.9**	**-3.28**	**-3.21**	**-3.07**	**-2.52**	-0.98	**-6.71**	-1.37
RIDGE	0.99	*2.28*	1.96	*2.44*	*2.57*	*2.15*	*5.57*	*5.91*
ELASTIC	-0.03	*2.04*	*2.03*	*2.33*	*2.8*	*2.65*	*5.86*	*6.17*
ADAP-ELASTIC	0	1.93	*1.97*	*2.3*	*2.76*	*2.71*	*5.8*	*6.23*
SCAD	0.4	1.04	1.41	1.91	1.87	*2.04*	*4.16*	*5.3*
BMS	**-4.48**	**-6.58**	**-7.29**	**-2.98**	**-4.07**	**-3.19**	-1.55	**-2.26**
BMA	**-6.21**	**-6.08**	**-5.92**	**-3.03**	**-3.09**	**-3.8**	-0.72	**-2.12**
CSR	0.27	0.98	1.02	0.8	0.67	1.11	1.15	*3.71*
NEURAL-NET	**-2.44**	*2.38*	1.87	1.51	*2.49*	*2.67*	*5.75*	*6.47*
RANDOM-FOREST	**-1.99**	-0.33	0.5	0.83	0.96	1.72	*1.98*	*5.05*
XG-BOOST	-1.74	1.4	1.72	*2.37*	1.73	1.68	*2.8*	*5.6*

Table 17: Conditional Diebold-Mariano statistics (8/5-digit headline)
Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

Given the size of the difference in the RMSE statistics, relative to the random walk, it is not surprising to note that in most cases the difference in forecasting performance is significantly different from zero, as is displayed in Table 17.

D.3 Four-digit data: core inflation

The out-of-sample RMSEs for the different models that employ the conditional forecasting function are contained in Table 18. In this case, we note that the official SARB forecasts continue to provide the most accurate forecast for the one-month and two-month horizon. This is largely due to the use of "off-model" information although the competing models also struggle to provide results that are equivalent to DIM over a one month horizon. Note also that the results from the "sparse" models are superior to those of the "dense" models, and for most of the sparse models, the results are particularly impressive over longer horizons, where the forecasting errors are about half the size of most benchmark models.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.19	0.35	0.46	0.52	0.71	1.04	1.05	1.1
AR-DIRECT	0.19	0.3	0.38	0.48	0.59	0.99	1.2	1.29
STOCH-VOL	0.18	0.28	0.34	0.41	0.51	0.83	1.01	1.11
RAND-WALK	0.18	0.26	0.32	0.39	0.47	0.76	0.94	1.04
BVAR	0.33	0.47	0.57	0.65	0.78	0.99	1.04	1.05
DIM	0.2	0.33	0.42	0.52	0.68	1.22	1.36	1.45
SARB	**0.13**	**0.24**	0.35	0.46	0.63	1.19	1.25	1.32
LINEAR	0.24	0.31	0.34	0.43	0.37	0.46	0.45	0.52
DFM-TF	0.23	0.34	0.43	0.52	0.61	0.95	1.05	1.11
DFM-3PRF	0.24	0.33	0.43	0.56	0.75	1.01	0.93	0.93
LASSO	0.27	0.28	0.28	0.29	0.26	0.33	0.35	0.37
LASSO-PSI	0.45	0.5	0.54	0.76	0.55	0.9	0.79	0.93
ADAP-LASSO	0.27	0.28	0.29	**0.29**	0.28	0.33	0.34	0.37
POST-LASSO	0.25	0.29	0.34	0.4	0.45	0.52	0.67	0.67
LASSO-ZERO	0.18	0.25	0.32	0.38	0.39	0.51	0.48	0.58
RIDGE	0.26	0.3	0.34	0.33	0.32	0.34	0.36	**0.34**
ELASTIC	0.25	0.26	0.29	0.3	**0.25**	0.31	0.33	0.35
ADAP-ELASTIC	0.26	0.27	0.31	0.31	0.26	0.31	0.34	0.34
SCAD	0.19	0.28	0.3	0.37	0.33	0.42	0.44	0.49
BMS	0.23	0.3	0.35	0.42	0.37	0.45	0.47	0.55
BMA	0.65	0.83	0.95	0.98	0.85	0.37	0.58	0.47
CSR	0.2	0.29	0.37	0.45	0.58	0.88	0.91	0.9
NEURAL-NET	0.28	0.29	0.32	0.31	0.31	0.49	0.6	0.49
RANDOM-FOREST	0.27	0.33	0.36	0.38	0.41	0.3	0.33	0.48
XG-BOOST	0.22	0.27	**0.28**	0.31	0.36	**0.24**	**0.26**	0.43

Table 18: Root-mean squared-error: four-digit core
See Table 2 for model acronyms.

Then lastly, when considering the results for the nonlinear statistical learning models we note, once again, that the XG boost model provides more accurate forecasts three, 12 and 18 steps ahead. Once again, over longer horizons, most of these models provide results that are significantly superior to the random walk model. However, over a 1-step-ahead horizon the forecasts from the random walk model are relatively impressive. These results are summarised by the Diebold and Mariano (1995) statistics, which are reported in Table 19.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	**-2.5**	**-2.77**	**-3.1**	**-2.71**	**-2.35**	-0.75	-0.22	-0.14
AR-DIRECT	**-2.62**	**-2.65**	**-2.55**	**-2.79**	**-2.05**	-0.98	-0.67	-0.49
STOCH-VOL	**-2.57**	**-2.44**	**-2.39**	-1.95	-1.72	-0.92	-0.51	-0.41
BVAR	**-3.96**	**-3.02**	**-2.77**	**-2.28**	-1.72	-0.65	-0.22	-0.03
DIM	-1.78	**-2.6**	**-2.91**	**-3.02**	**-3.33**	**-2.21**	-1.25	-0.89
SARB	1.37	0.49	-0.68	-1.49	-1.82	-1.84	-0.84	-0.56
LINEAR	**-2.87**	-1.34	-0.42	-0.87	1.18	1.61	*5.95*	*7.28*
DFM-TF	**-3.59**	**-2.55**	**-2.01**	-1.69	-1.21	-0.74	-0.25	-0.19
DFM-3PRF	**-3.15**	**-2.19**	-1.82	-1.76	-1.84	**-3.76**	0.04	1.81
LASSO	**-3.86**	-0.43	0.96	*2.14*	*1.99*	*2.04*	*4.6*	*8.69*
LASSO-PSI	**-6.06**	**-4.97**	**-2.58**	**-2.25**	**-2.89**	-1.24	1.61	0.27
ADAP-LASSO	**-3.74**	-0.44	0.68	*2.11*	1.84	*2.06*	*4.79*	*8.58*
POST-LASSO	**-2.9**	-0.72	-0.52	-0.26	0.23	1.29	1.23	*6.08*
LASSO-ZERO	-0.97	0.33	-0.01	0.1	1.24	1.34	*5.66*	*6.02*
RIDGE	**-2.82**	-0.73	-0.31	0.82	1.33	*2.39*	*9.17*	*8.98*
ELASTIC	**-3.04**	-0.01	0.65	1.68	*2.01*	*2.21*	*5.4*	*8.85*
ADAP-ELASTIC	**-3.22**	-0.24	0.25	1.45	1.93	*2.21*	*5.59*	*8.86*
SCAD	-1.84	-1.78	0.65	0.39	1.58	1.58	*4.9*	*7.14*
BMS	**-2.74**	-1.32	-1.2	-0.79	1.56	1.68	*4.85*	*6.87*
BMA	**-9.97**	**-6.37**	**-3.66**	**-3.59**	**-2.73**	1.81	*2.19*	*6.7*
CSR	**-2.37**	-1.19	-1.66	**-2.29**	**-2.4**	-0.52	0.11	0.66
NEURAL-NET	**-2.96**	-0.95	-0.05	1.09	1.31	1.86	*2.39*	*6.16*
RANDOM-FOREST	**-3.17**	-1.9	-1.34	0.16	0.92	*2.16*	*5.29*	*7.79*
XG-BOOST	**-2.47**	-0.31	1.27	*2.24*	1.15	*2.22*	*6.84*	*7.67*

Table 19: Conditional Diebold-Mariano statistics (four-digit core)

Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

D.4 Eight/five-digit data: core inflation

Table 20 contains the out-of-sample RMSE for the different models, when we employ a conditional forecasting function. Once again, the results for all of the models that employ variable selection techniques are superior to all the benchmark models when the horizon is greater than two months. In addition, note that the RMSEs for these models are slightly better than what was provided when we make use of the four-digit data in certain cases.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.18	0.34	0.45	0.53	0.75	1.15	1.35	1.45
AR-DIRECT	0.19	0.29	0.36	0.45	0.55	0.89	1.06	1.12
STOCH-VOL	0.18	0.27	0.34	0.41	0.5	0.8	0.98	1.07
RAND-WALK	0.18	0.26	0.32	0.39	0.47	0.76	0.94	1.04
BVAR	0.3	0.42	0.51	0.59	0.72	1.01	1.13	1.22
DIM	0.2	0.33	0.42	0.52	0.68	1.22	1.36	1.45
SARB	**0.13**	**0.24**	0.35	0.46	0.63	1.19	1.25	1.32
DFM-TF	0.18	0.33	0.44	0.52	0.76	1.07	1.13	1.12
DFM-3PRF	0.27	0.44	0.6	0.75	1.01	1.08	1.11	1.32
LASSO	0.28	0.25	0.23	0.23	0.26	0.32	0.25	0.32
LASSO-PSI	0.56	0.73	0.73	0.55	0.81	0.58	0.84	0.79
ADAP-LASSO	0.27	0.25	**0.23**	**0.23**	0.28	0.32	0.26	0.32
POST-LASSO	0.22	0.29	0.3	0.31	0.36	0.34	0.4	0.44
LASSO-ZERO	0.92	1.11	0.77	1.06	1.3	1.53	1.54	1.62
RIDGE	0.27	0.31	0.32	0.25	0.25	0.29	0.27	0.3
ELASTIC	0.26	0.24	0.23	0.24	0.25	0.33	**0.24**	0.32
ADAP-ELASTIC	0.27	0.25	0.24	0.24	**0.24**	0.32	0.24	0.32
SCAD	0.19	0.27	0.32	0.33	0.4	0.33	0.39	0.38
BMS	0.73	1	1.12	1.16	1.2	1.61	2.15	1.41
BMA	0.28	0.3	0.37	0.35	0.36	0.93	0.52	1.22
CSR	0.2	0.29	0.36	0.4	0.43	0.55	0.62	0.62
NEURAL-NET	0.22	0.28	0.25	0.28	0.3	0.29	0.32	**0.3**
RANDOM-FOREST	0.25	0.28	0.29	0.3	0.33	0.31	0.43	0.5
XG-BOOST	0.22	0.25	0.27	0.27	0.29	**0.28**	0.39	0.46

Table 20: Root-mean squared-error: 8/5-digit core
See Table 2 for model acronyms.

The results of the nonlinear statistical learning models are also similar to what was reported for the four-digit data, but once again, they are slightly more impressive, where the neural network and boosting models provide some of the most accurate forecasts for horizons that are equal to 12 and 24 months ahead.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	**-1.99**	**-2.61**	**-2.66**	**-2.35**	**-2.86**	-1.91	-1.5	-1.19
AR-DIRECT	**-2.17**	**-2.05**	-1.93	**-2.03**	-1.41	-0.63	-0.36	-0.21
STOCH-VOL	**-2.65**	**-2.36**	**-2.28**	-1.77	-1.46	-0.62	-0.31	-0.18
BVAR	**-3.09**	**-2.78**	**-2.7**	**-2.31**	-1.95	-0.92	-0.44	-0.34
DIM	-1.78	**-2.6**	**-2.91**	**-3.02**	**-3.33**	**-2.21**	-1.25	-0.89
SARB	1.37	0.49	-0.68	-1.49	-1.82	-1.84	-0.84	-0.56
DFM-TF	-0.79	-1.59	-1.84	-1.55	-1.73	**-1.99**	-1.45	-0.58
DFM-3PRF	**-2.87**	**-2.19**	**-2.04**	-1.95	**-2.3**	-1.84	**-4.17**	**-6.9**
LASSO	**-3.46**	0.23	1.96	*2.48*	*1.99*	*2.03*	*7.45*	*9.23*
LASSO-PSI	**-4.76**	**-4.95**	**-3.07**	**-2.13**	-1.22	0.91	1.17	0.55
ADAP-LASSO	**-3.09**	0.3	*2.07*	*2.45*	*1.98*	*1.99*	*7.89*	*9.21*
POST-LASSO	**-2.36**	-0.61	0.42	1.87	*1.99*	*2.01*	*3.75*	*8.12*
LASSO-ZERO	**-9.4**	**-3.47**	**-3.11**	**-2.66**	**-2.06**	**-2.13**	**-3.12**	-1.06
RIDGE	**-3.3**	-1.21	-0.06	*2.05*	*2.15*	*2.18*	*5.98*	*8.73*
ELASTIC	**-2.98**	0.5	1.69	*2.21*	*2.02*	1.94	*7.19*	*9.21*
ADAP-ELASTIC	**-2.94**	0.34	1.63	*2.17*	*2.08*	*1.98*	*7.36*	*9.16*
SCAD	-1.81	-0.55	0.03	0.9	0.74	*2.04*	*4.66*	*8.14*
BMS	**-7.85**	**-4.25**	**-3.25**	**-4.01**	**-2.11**	-1.1	**-3.84**	-0.92
BMA	**-2.16**	-0.8	-1.1	0.58	0.88	-0.46	*4.69*	-0.36
CSR	**-2.38**	-1.12	-1.1	-0.22	0.98	1.1	1.78	*6.1*
NEURAL-NET	**-2.07**	-0.39	1.87	1.8	1.44	*2*	*6.38*	*8.6*
RANDOM-FOREST	**-2.59**	-0.58	1.32	*2.5*	1.74	*2.13*	*2.77*	*8.09*
XG-BOOST	**-2.74**	0.54	1.95	*2.14*	1.94	*2.17*	*3.46*	*8.7*

Table 21: Conditional Diebold-Mariano statistics (8/5-digit core)

Boldface indicates that the difference in forecasting accuracy is significantly different from zero in favour of the random-walk forecast, while *italics* indicate that the difference is significant, but in favour of the competing model. See Table 2 for model acronyms.

In terms of whether or not the difference in forecasting performance is significantly different from zero, the Diebold and Mariano (1995) statistics suggest the difference in forecasting performance is mostly significantly different from zero, in favour of the random walk model over 1-step-ahead horizon. However, over longer horizons, most of the competing models provide a significant improvement over the random walk model. These results are summarised in Table 21.

E Mean absolute percentage error

E.1 Four-digit data: headline inflation

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.35	0.54	0.63	0.73	0.94	1.38	1.48	1.51
AR-DIRECT	0.34	0.52	0.6	0.69	0.88	1.4	1.44	1.48
STOCH-VOL	0.32	0.49	0.56	0.65	0.76	1.01	1.15	**1.17**
RAND-WALK	0.32	0.49	0.56	0.64	0.74	**0.98**	1.13	1.25
BVAR	0.45	0.63	0.77	0.88	1.03	1.21	1.24	1.24
DIM	0.24	0.41	0.52	0.64	0.79	1.34	1.44	1.53
SARB	**0.13**	**0.21**	**0.32**	**0.44**	**0.62**	1.27	1.41	1.47
LINEAR	0.38	0.63	0.94	0.93	1.2	2.01	2.49	2.8
DFM-TF	0.34	0.51	0.65	0.8	1.03	1.65	1.63	1.55
DFM-3PRF	0.34	0.52	0.65	0.78	1.07	1.52	1.48	1.83
LASSO	0.32	0.63	0.77	0.88	1.32	1.78	2.37	1.95
LASSO-PSI	0.58	1.37	1.19	1.07	1.43	3.01	2.59	1.99
ADAP-LASSO	0.32	0.63	0.83	0.93	1.35	1.77	2.15	1.85
POST-LASSO	0.34	0.5	0.52	0.62	0.87	1.49	1.61	2.13
LASSO-ZERO	0.33	0.55	0.71	0.9	0.93	2.29	1.75	2.33
RIDGE	0.38	0.77	0.91	1.08	1.43	1.88	2.73	2.11
ELASTIC	0.33	0.66	0.86	0.95	1.43	1.68	2.51	2.06
ADAP-ELASTIC	0.33	0.66	0.92	1.01	1.46	1.72	2.44	2.01
SCAD	0.34	0.67	0.68	0.86	1.09	1.87	1.7	2.55
BMS	0.35	0.58	0.74	0.81	0.86	2.29	1.31	2.43
BMA	2.36	1.37	0.71	1.46	1.75	3.76	**1.1**	1.94
CSR	0.33	0.49	0.52	0.59	0.72	1.52	1.36	1.99
NEURAL-NET	0.35	0.6	0.79	0.83	1.05	1.62	1.38	1.61
RANDOM-FOREST	0.36	0.53	0.7	0.87	1.06	1.24	1.4	1.47
XG-BOOST	0.35	0.5	0.57	0.76	0.95	1.07	1.11	1.34

Table 22: Unconditional mean absolute percentage error (four-digit headline inflation)
See Table 2 for model acronyms.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.35	0.54	0.63	0.73	0.94	1.38	1.48	1.51
AR-DIRECT	0.34	0.57	0.72	0.83	0.99	1.3	1.32	1.4
STOCH-VOL	0.32	0.49	0.56	0.65	0.76	1.01	1.15	1.17
RAND-WALK	0.32	0.49	0.56	0.64	0.74	0.98	1.13	1.25
BVAR	0.45	0.63	0.77	0.88	1.03	1.21	1.24	1.24
DIM	0.37	0.54	0.62	0.72	0.88	1.35	1.45	1.53
SARB	**0.13**	**0.21**	**0.32**	0.44	0.62	1.27	1.41	1.47
LINEAR	0.38	0.49	0.55	0.5	0.55	0.51	0.6	0.54
DFM-TF	0.34	0.51	0.63	0.74	0.89	1.38	1.31	1.19
DFM-3PRF	0.34	0.5	0.58	0.63	0.78	0.99	1.13	1.29
LASSO	0.33	0.41	0.46	0.44	0.51	**0.42**	0.45	0.4
LASSO-PSI	0.52	1.04	1	0.91	1.05	1.42	1.79	2.28
ADAP-LASSO	0.32	0.42	0.49	0.45	0.54	0.44	0.44	0.4
POST-LASSO	0.34	0.48	0.45	0.51	0.6	0.76	0.64	0.78
LASSO-ZERO	0.33	0.48	0.52	0.52	0.5	0.5	0.46	0.51
RIDGE	0.37	0.49	0.49	0.47	0.5	0.44	0.51	0.4
ELASTIC	0.34	0.42	0.48	0.44	0.52	0.43	0.46	**0.39**
ADAP-ELASTIC	0.32	0.42	0.5	0.46	0.53	0.45	0.45	0.4
SCAD	0.35	0.51	0.45	0.47	0.46	0.57	0.51	0.6
BMS	0.35	0.49	0.53	0.48	0.49	0.51	0.47	0.52
BMA	0.33	0.52	0.5	0.5	0.5	0.5	0.46	0.49
CSR	0.33	0.47	0.5	0.49	0.53	0.91	0.98	0.97
NEURAL-NET	0.41	0.39	0.41	0.41	**0.4**	0.43	0.48	0.5
RANDOM-FOREST	0.35	0.43	0.49	0.52	0.57	0.53	0.47	0.57
XG-BOOST	0.34	0.39	0.37	**0.4**	0.44	0.45	**0.36**	0.5

Table 23: Conditional mean absolute percentage error (four-digit headline inflation)
See Table 2 for model acronyms.

E.2 Eight/five-digit data: headline inflation

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.34	0.53	0.6	0.7	0.89	1.21	1.25	1.26
AR-DIRECT	0.34	0.51	0.6	0.71	0.92	1.48	1.54	1.47
STOCH-VOL	0.32	0.49	0.56	0.66	0.78	1.03	1.18	**1.21**
RAND-WALK	0.32	0.49	0.56	0.64	0.74	**0.98**	**1.13**	1.25
BVAR	0.37	0.51	0.63	0.75	0.96	1.28	1.44	1.56
DIM	0.24	0.41	0.52	0.64	0.79	1.34	1.44	1.53
SARB	**0.13**	**0.21**	**0.32**	**0.44**	**0.62**	1.27	1.41	1.47
DFM-TF	0.33	0.49	0.61	0.77	1.09	1.68	1.46	1.53
DFM-3PRF	0.34	0.53	0.68	0.82	1.13	1.63	1.76	2.3
LASSO	0.34	0.46	0.46	0.6	0.93	1.51	1.51	1.7
LASSO-PSI	1.2	0.96	1.76	1.63	1.35	2.1	1.94	1.63
ADAP-LASSO	0.34	0.46	0.46	0.62	0.92	1.53	1.51	1.72
POST-LASSO	0.35	0.58	0.72	0.84	1.08	1.4	1.68	1.86
LASSO-ZERO	0.42	1.2	1.95	2.05	4.34	2.42	1.88	1.57
RIDGE	0.3	0.45	0.55	0.63	0.92	1.82	1.6	2.36
ELASTIC	0.35	0.44	0.51	0.6	0.87	1.4	1.57	1.87
ADAP-ELASTIC	0.35	0.42	0.51	0.61	0.88	1.41	1.66	1.88
SCAD	0.33	0.58	0.73	0.8	1.16	1.74	1.6	2
BMS	1.16	1.89	1.9	2.03	1.71	3.29	2.47	1.88
BMA	1.82	1.45	2.13	2.42	1.92	4.44	2.4	2.23
CSR	0.34	0.52	0.66	0.84	1.03	1.4	1.6	2.09
NEURAL-NET	0.4	0.51	0.59	0.83	0.91	1.26	1.2	1.22
RANDOM-FOREST	0.43	0.66	0.9	1.07	1.18	1.29	1.42	1.52
XG-BOOST	0.39	0.57	0.83	0.98	1.23	1.33	1.46	1.56

Table 24: Unconditional mean absolute percentage error (8/5-digit headline inflation)
See Table 2 for model acronyms.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.34	0.53	0.6	0.7	0.89	1.21	1.25	1.26
AR-DIRECT	0.34	0.57	0.72	0.84	1.04	1.39	1.46	1.41
STOCH-VOL	0.32	0.49	0.56	0.66	0.78	1.03	1.18	1.21
RAND-WALK	0.32	0.49	0.56	0.64	0.74	0.98	1.13	1.25
BVAR	0.37	0.51	0.63	0.75	0.96	1.28	1.44	1.56
DIM	0.24	0.41	0.52	0.64	0.79	1.34	1.44	1.53
SARB	**0.13**	**0.21**	0.32	0.44	0.62	1.27	1.41	1.47
DFM-TF	0.33	0.53	0.66	0.84	1.16	1.47	1.39	1.29
DFM-3PRF	0.34	0.54	0.66	0.75	1	1.12	1.5	1.86
LASSO	0.34	0.35	0.33	0.38	0.36	0.32	0.46	0.43
LASSO-PSI	1.01	0.92	1.49	1.61	1.54	2.1	1.71	1.88
ADAP-LASSO	0.34	0.36	0.33	0.39	0.37	**0.31**	0.46	0.44
POST-LASSO	0.34	0.48	0.46	0.5	0.56	0.55	0.57	0.59
LASSO-ZERO	0.42	1.19	1.86	1.91	3.49	1.17	2.39	2.29
RIDGE	0.3	0.29	0.34	**0.32**	**0.33**	0.43	0.39	0.49
ELASTIC	0.34	0.34	**0.31**	0.37	0.35	0.33	0.41	0.44
ADAP-ELASTIC	0.35	0.35	0.32	0.39	0.36	0.32	0.43	0.45
SCAD	0.33	0.43	0.48	0.51	0.5	0.5	0.46	0.6
BMS	1.36	2.16	1.55	2.27	1.31	1.67	1.39	2.1
BMA	1.84	1.73	1.63	2.64	1.49	2.43	1.28	2.79
CSR	0.34	0.48	0.54	0.59	0.62	0.69	0.78	0.9
NEURAL-NET	0.39	0.34	0.41	0.48	0.4	0.37	**0.38**	**0.38**
RANDOM-FOREST	0.43	0.53	0.61	0.64	0.62	0.51	0.6	0.61
XG-BOOST	0.38	0.41	0.48	0.53	0.5	0.49	0.53	0.5

Table 25: Conditional mean absolute percentage error (8/5-digit headline inflation)

See Table 2 for model acronyms.

E.3 Four-digit data: core inflation

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.14	0.23	0.3	0.38	0.49	0.89	1.1	1.18
AR-DIRECT	0.14	0.22	0.31	0.39	0.47	0.81	0.96	1.04
STOCH-VOL	0.13	0.2	0.27	0.33	0.42	0.73	0.91	1.06
RAND-WALK	0.12	**0.19**	**0.25**	**0.31**	**0.38**	0.65	0.81	0.96
BVAR	0.26	0.38	0.47	0.56	0.7	0.86	0.89	0.9
DIM	0.15	0.26	0.35	0.44	0.62	1.18	1.3	1.38
SARB	**0.11**	0.2	0.29	0.4	0.56	1.14	1.17	1.21
LINEAR	0.19	0.31	0.42	0.55	0.58	2.08	1.64	1.21
DFM-TF	0.18	0.29	0.4	0.52	0.7	1.34	1.52	1.73
DFM-3PRF	0.18	0.28	0.36	0.49	0.78	1.42	1.31	1.3
LASSO	0.21	0.29	0.34	0.38	0.55	1.3	1.16	1.18
LASSO-PSI	0.38	0.48	0.53	0.83	0.54	1.03	1.04	1.4
ADAP-LASSO	0.21	0.3	0.35	0.39	0.58	1.32	1.11	1.17
POST-LASSO	0.19	0.27	0.36	0.46	0.57	1.25	0.88	1.42
LASSO-ZERO	0.14	0.22	0.33	0.46	0.52	1.47	1.37	2.82
RIDGE	0.21	0.33	0.5	0.58	0.54	1.61	1.16	**0.83**
ELASTIC	0.2	0.3	0.36	0.38	0.51	1.31	1.01	1.09
ADAP-ELASTIC	0.21	0.32	0.36	0.38	0.52	1.29	0.98	1.12
SCAD	0.13	0.24	0.27	0.32	0.56	0.91	1.11	1.38
BMS	0.18	0.29	0.39	0.5	0.53	1.68	1.35	2.51
BMA	0.61	0.75	0.78	0.76	0.59	1.33	**0.67**	2.47
CSR	0.15	0.25	0.35	0.44	0.65	1.23	1.23	1.2
NEURAL-NET	0.24	0.33	0.38	0.45	0.65	1	1.02	1.02
RANDOM-FOREST	0.19	0.32	0.45	0.56	0.71	**0.6**	0.75	1.08
XG-BOOST	0.17	0.29	0.36	0.49	0.66	0.75	0.73	1.19

Table 26: Unconditional mean absolute percentage error (four-digit core inflation)
See Table 2 for model acronyms.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.14	0.23	0.3	0.38	0.49	0.89	1.1	1.18
AR-DIRECT	0.14	0.27	0.37	0.44	0.6	0.86	0.85	0.95
STOCH-VOL	0.13	0.2	0.27	0.33	0.42	0.73	0.91	1.06
RAND-WALK	0.12	**0.19**	0.25	0.31	0.38	0.65	0.81	0.96
BVAR	0.26	0.38	0.47	0.56	0.7	0.86	0.89	0.9
DIM	0.15	0.26	0.35	0.44	0.62	1.18	1.3	1.38
SARB	**0.11**	0.2	0.29	0.4	0.56	1.14	1.17	1.21
LINEAR	0.19	0.24	0.27	0.35	0.28	0.37	0.35	0.41
DFM-TF	0.18	0.28	0.35	0.43	0.52	0.81	0.9	0.95
DFM-3PRF	0.18	0.27	0.35	0.45	0.62	0.88	0.81	0.82
LASSO	0.21	0.22	0.23	0.23	0.21	0.26	0.27	0.29
LASSO-PSI	0.38	0.42	0.43	0.62	0.42	0.72	0.63	0.71
ADAP-LASSO	0.21	0.23	0.23	0.23	0.22	0.25	0.26	0.29
POST-LASSO	0.19	0.23	0.29	0.33	0.33	0.46	0.53	0.57
LASSO-ZERO	0.14	0.2	0.26	0.32	0.3	0.43	0.41	0.46
RIDGE	0.21	0.24	0.27	0.28	0.26	0.25	0.27	**0.25**
ELASTIC	0.2	0.21	0.23	**0.22**	**0.2**	0.24	0.25	0.27
ADAP-ELASTIC	0.2	0.22	0.24	0.24	0.2	0.23	0.25	0.26
SCAD	0.13	0.22	0.24	0.29	0.27	0.32	0.35	0.38
BMS	0.18	0.24	0.29	0.33	0.29	0.37	0.39	0.43
BMA	0.61	0.79	0.88	0.89	0.8	0.29	0.5	0.38
CSR	0.15	0.23	0.31	0.39	0.52	0.79	0.82	0.8
NEURAL-NET	0.21	0.23	0.26	0.25	0.25	0.37	0.37	0.35
RANDOM-FOREST	0.19	0.26	0.3	0.32	0.35	0.22	0.24	0.42
XG-BOOST	0.16	0.2	**0.2**	0.24	0.29	**0.19**	**0.2**	0.34

Table 27: Conditional mean absolute percentage error (four-digit core inflation)
See Table 2 for model acronyms.

E.4 Eight/five-digit data: core inflation

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.13	0.21	0.28	0.35	0.45	0.79	0.95	1.01
AR-DIRECT	0.13	0.21	0.29	0.36	0.45	0.9	1.18	1.35
STOCH-VOL	0.13	0.2	0.27	0.33	0.41	0.7	0.88	1.01
RAND-WALK	0.12	**0.19**	**0.25**	**0.31**	**0.38**	**0.65**	0.81	0.96
BVAR	0.22	0.32	0.41	0.5	0.64	0.91	1.03	1.14
DIM	0.15	0.26	0.35	0.44	0.62	1.18	1.3	1.38
SARB	**0.11**	0.2	0.29	0.4	0.56	1.14	1.17	1.21
DFM-TF	0.14	0.21	0.27	0.33	0.41	0.89	1.22	1.38
DFM-3PRF	0.2	0.33	0.46	0.59	0.85	1.4	1.33	1.52
LASSO	0.22	0.28	0.31	0.46	0.76	1.35	0.86	1.31
LASSO-PSI	0.43	0.64	0.78	0.54	0.7	0.87	1.51	1.14
ADAP-LASSO	0.21	0.28	0.3	0.42	0.75	1.35	0.84	1.31
POST-LASSO	0.17	0.28	0.38	0.45	0.6	1.08	1.18	1.37
LASSO-ZERO	0.83	0.95	0.69	0.88	1.16	1.78	1.97	1.36
RIDGE	0.22	0.37	0.46	0.43	0.53	0.93	1.08	1.4
ELASTIC	0.21	0.27	0.3	0.5	0.63	1.32	0.84	1.29
ADAP-ELASTIC	0.21	0.27	0.29	0.47	0.61	1.28	0.82	1.29
SCAD	0.14	0.24	0.33	0.47	0.66	1	1.14	1.45
BMS	0.65	0.9	0.97	1.09	0.87	1.83	1.81	1.5
BMA	3.2	3.18	3.07	2.68	2.87	6.15	26.21	8.91
CSR	0.16	0.27	0.38	0.44	0.52	0.92	1.18	1.19
NEURAL-NET	0.19	0.31	0.32	0.34	0.46	0.72	**0.69**	**0.89**
RANDOM-FOREST	0.18	0.27	0.34	0.44	0.57	0.83	1.07	1.27
XG-BOOST	0.17	0.27	0.31	0.4	0.52	0.79	1.05	1.27

Table 28: Unconditional mean absolute percentage error (8/5-digit core inflation)
See Table 2 for model acronyms.

Model	1-step	2-step	3-step	4-step	6-step	12-step	18-step	24-step
AR(1)	0.13	0.21	0.28	0.35	0.45	0.79	0.95	1.01
AR-DIRECT	0.13	0.26	0.36	0.44	0.65	1.05	1.3	1.4
STOCH-VOL	0.13	0.2	0.27	0.33	0.41	0.7	0.88	1.01
RAND-WALK	0.12	0.19	0.25	0.31	0.38	0.65	0.81	0.96
BVAR	0.22	0.32	0.41	0.5	0.64	0.91	1.03	1.14
DIM	0.15	0.26	0.35	0.44	0.62	1.18	1.3	1.38
SARB	**0.11**	0.2	0.29	0.4	0.56	1.14	1.17	1.21
DFM-TF	0.14	0.27	0.37	0.44	0.62	0.94	1.03	1.07
DFM-3PRF	0.2	0.34	0.46	0.59	0.83	0.95	1.09	1.29
LASSO	0.22	0.21	0.18	0.18	0.22	0.22	0.19	0.24
LASSO-PSI	0.45	0.59	0.61	0.48	0.58	0.44	0.7	0.53
ADAP-LASSO	0.21	0.2	**0.17**	**0.18**	0.22	0.23	0.2	0.24
POST-LASSO	0.17	0.22	0.24	0.25	0.28	0.27	0.29	0.34
LASSO-ZERO	0.83	0.95	0.67	0.86	1.08	1.32	1.22	1.23
RIDGE	0.21	0.25	0.26	0.2	0.2	0.21	0.21	**0.21**
ELASTIC	0.2	0.19	0.18	0.19	**0.2**	0.22	**0.19**	0.24
ADAP-ELASTIC	0.21	0.19	0.19	0.2	0.2	0.22	0.19	0.24
SCAD	0.14	0.21	0.27	0.27	0.28	0.26	0.29	0.31
BMS	0.65	0.86	0.91	1.04	0.98	1.23	1.77	1.24
BMA	0.19	0.23	0.31	0.27	0.28	0.58	0.41	0.81
CSR	0.16	0.24	0.31	0.34	0.34	0.46	0.53	0.51
NEURAL-NET	0.16	0.21	0.21	0.22	0.23	0.24	0.25	0.24
RANDOM-FOREST	0.18	0.21	0.23	0.23	0.26	0.25	0.33	0.44
XG-BOOST	0.16	**0.18**	0.21	0.21	0.24	**0.19**	0.28	0.35

Table 29: Conditional mean absolute percentage error (8/5-digit core inflation)

See Table 2 for model acronyms.

F Disaggregated data at 8/5-digit level

Figure 6: Consumer Price Index - 8/5-digit data

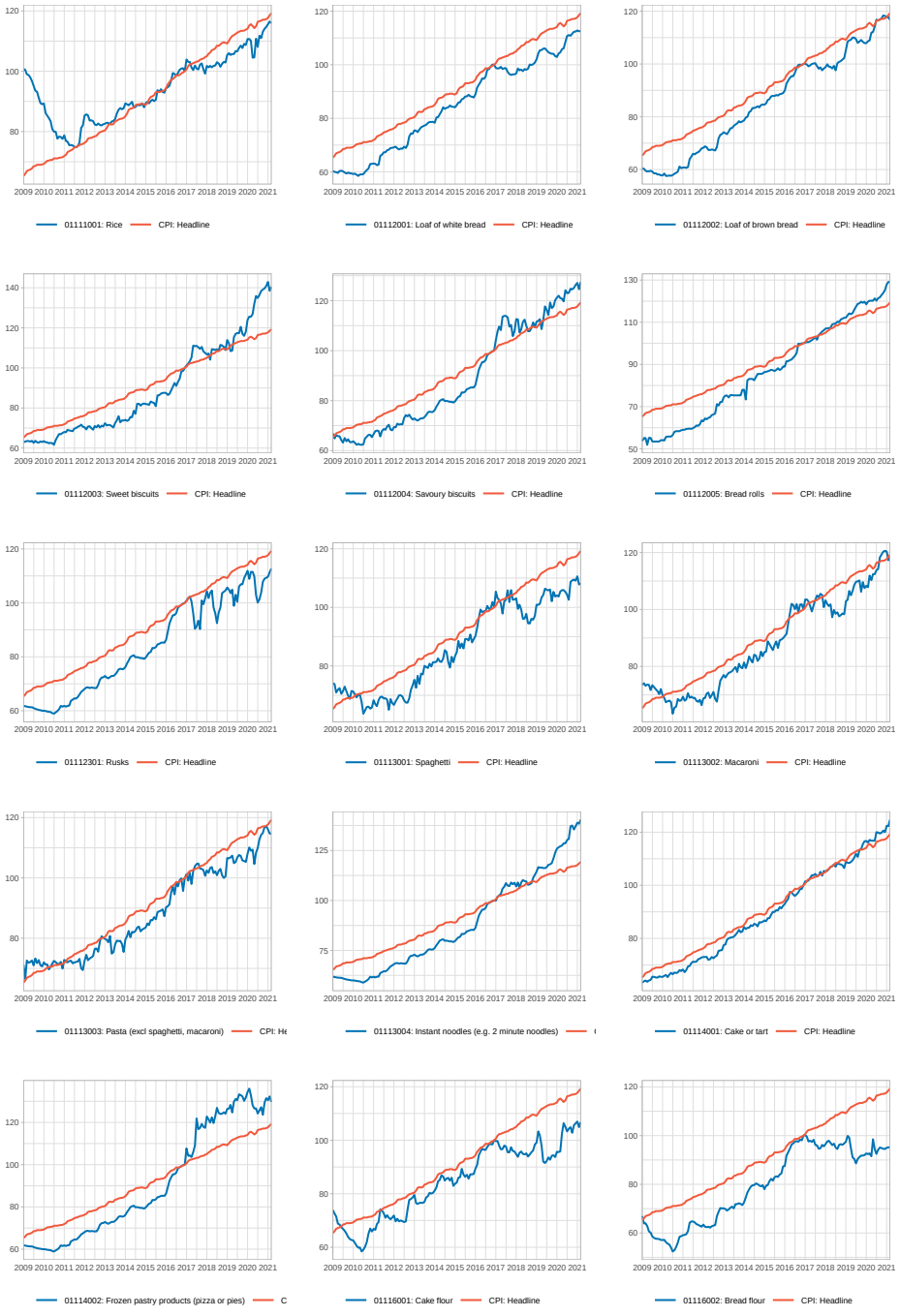


Figure 7: Consumer Price Index - 8/5-digit data

Figure 8: Consumer Price Index - 8/5-digit data

Figure 9: Consumer Price Index - 8/5-digit data

Figure 10: Consumer Price Index - 8/5-digit data

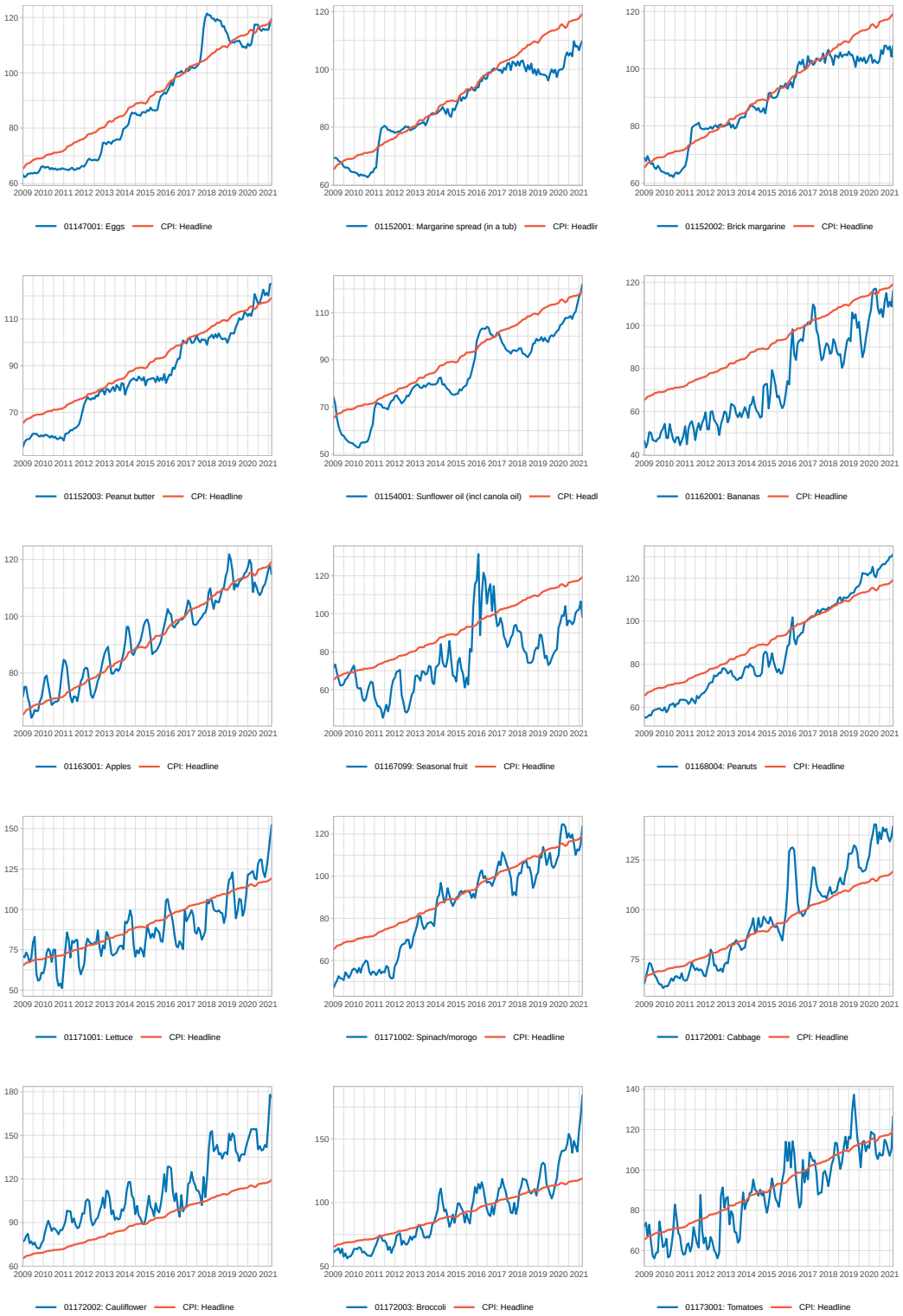


Figure 11: Consumer Price Index - 8/5-digit data

Figure 12: Consumer Price Index - 8/5-digit data

Figure 13: Consumer Price Index - 8/5-digit data

Figure 14: Consumer Price Index - 8/5-digit data

Figure 15: Consumer Price Index - 8/5-digit data

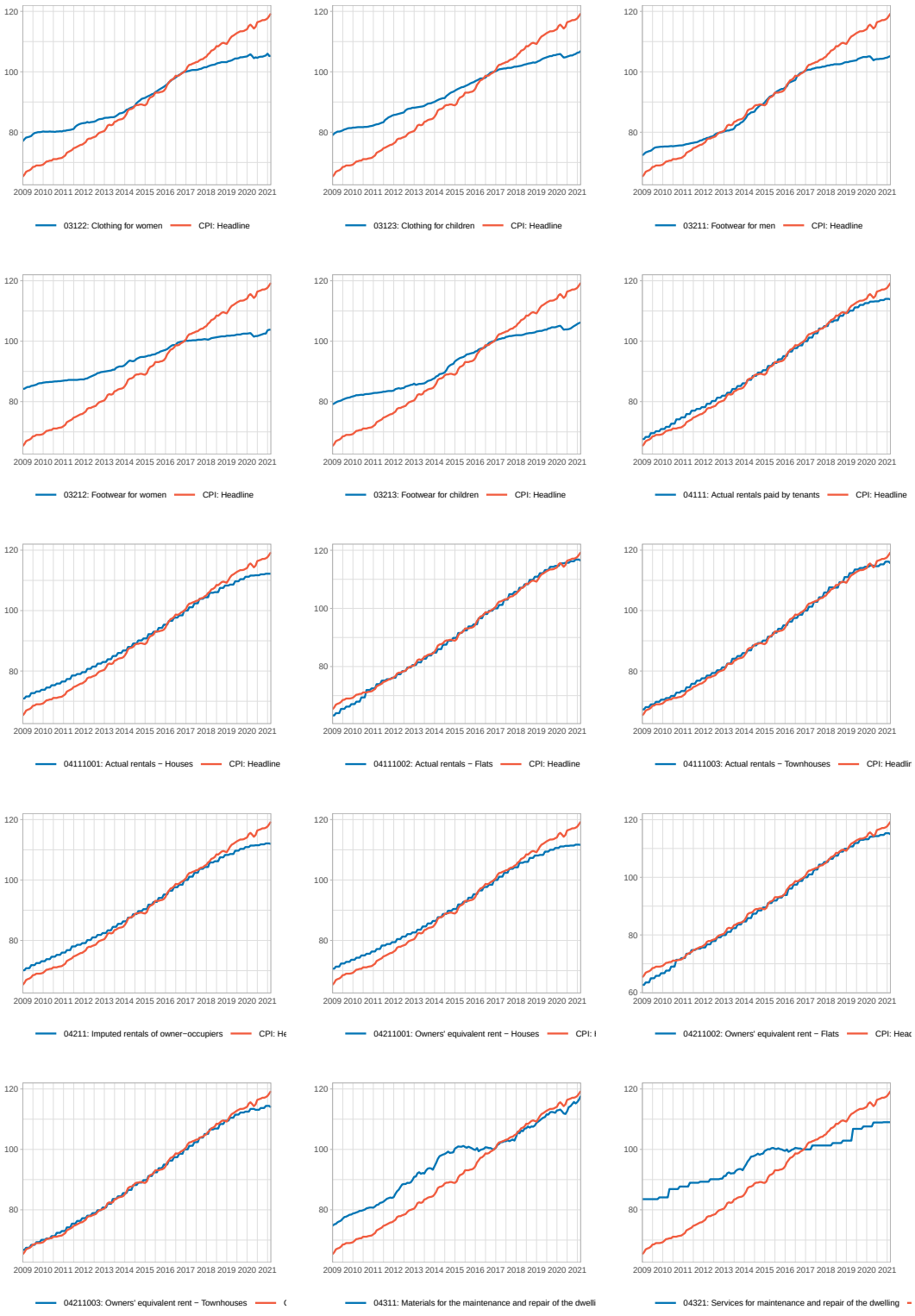


Figure 16: Consumer Price Index - 8/5-digit data

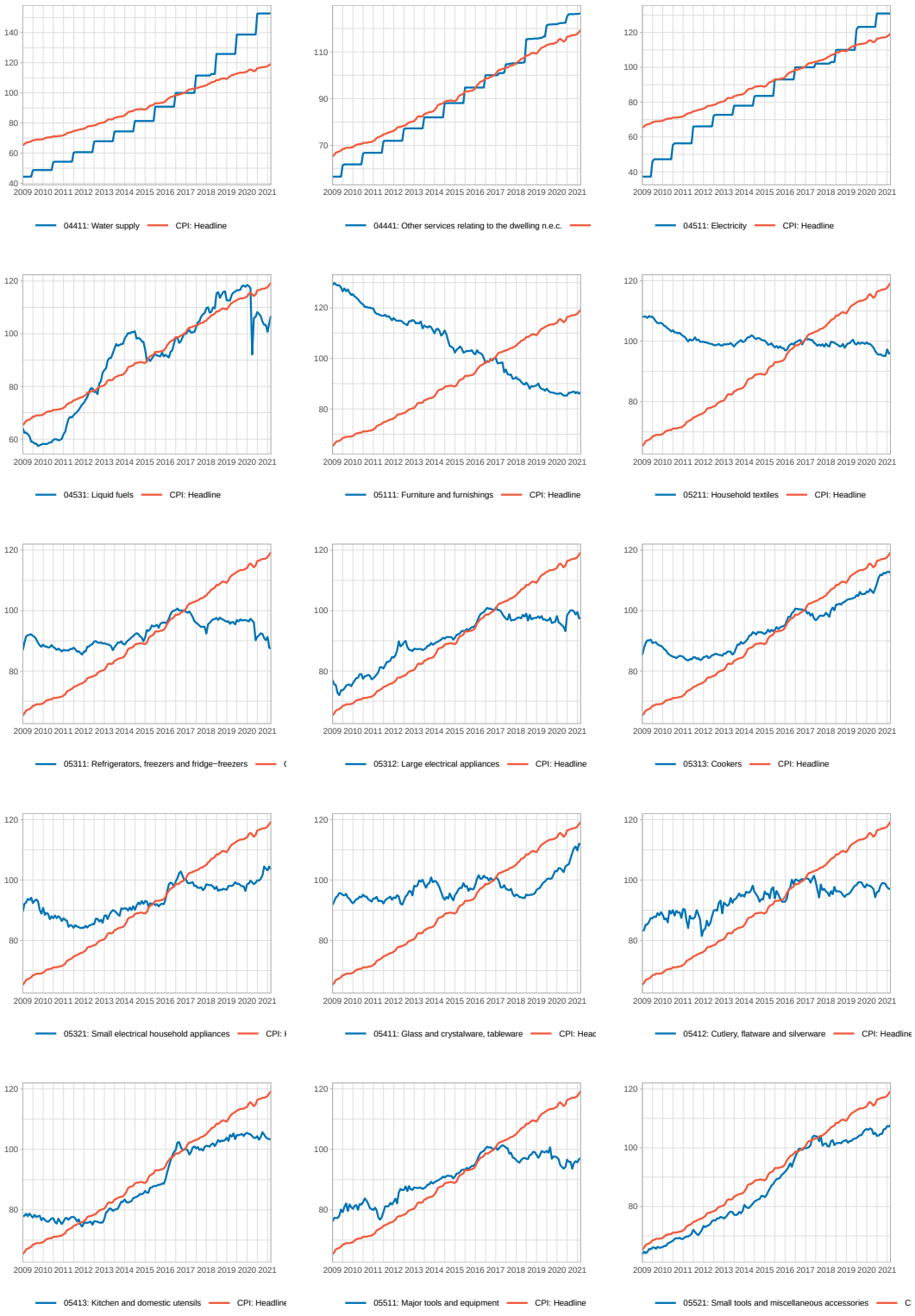


Figure 17: Consumer Price Index - 8/5-digit data

Figure 18: Consumer Price Index - 8/5-digit data

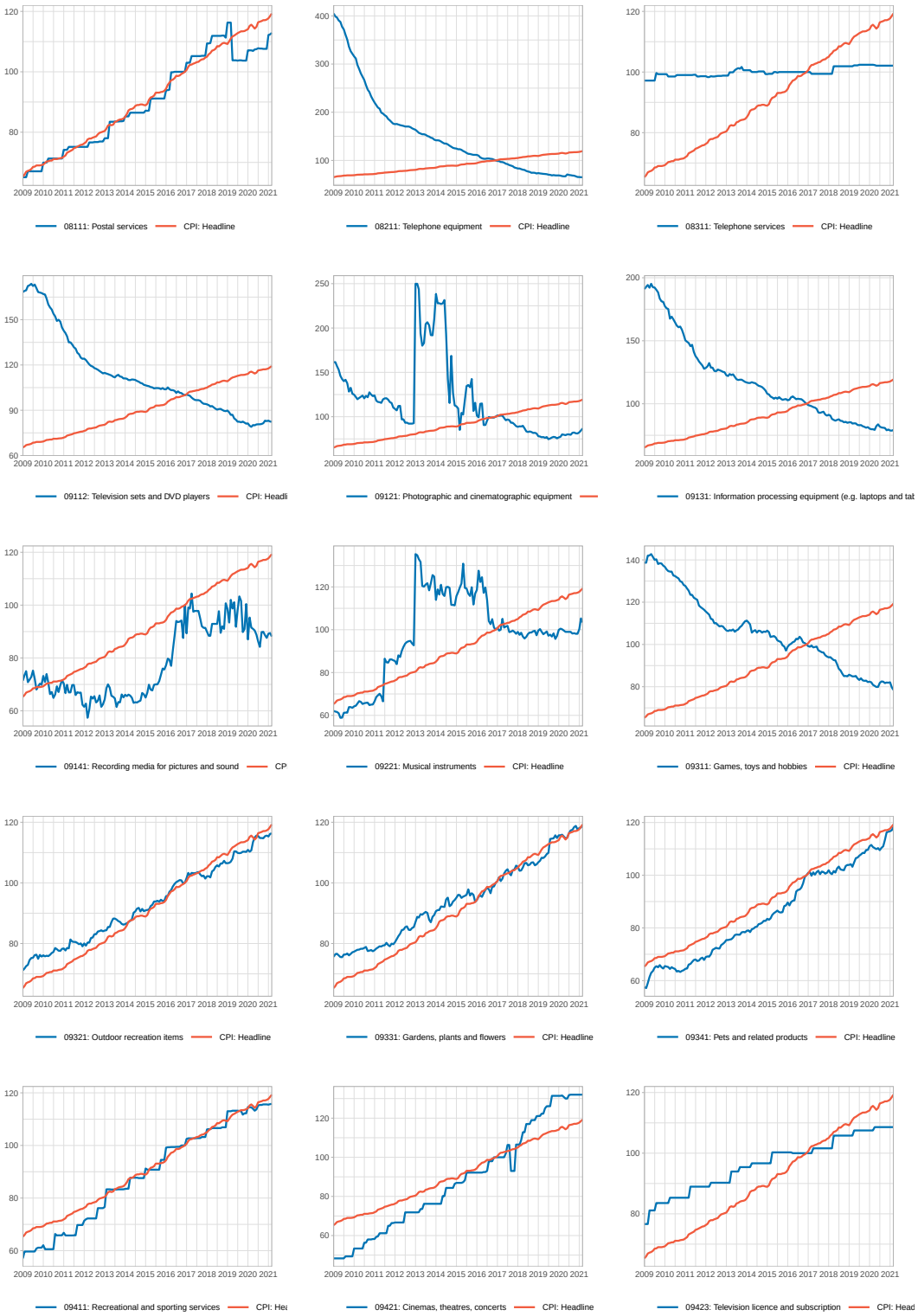


Figure 19: Consumer Price Index - 8/5-digit data

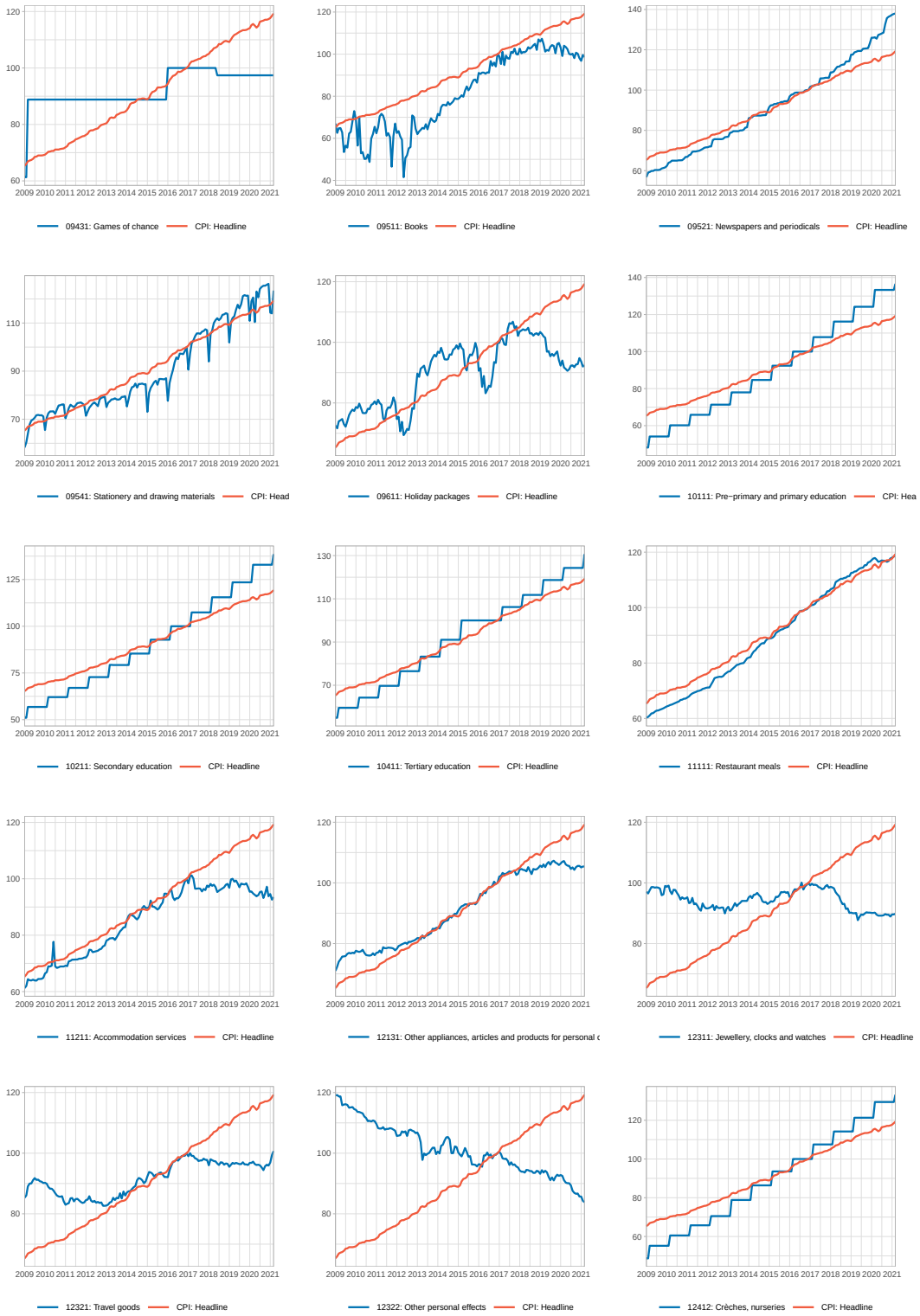


Figure 20: Consumer Price Index - 8/5-digit data