UNIVERSITY OF BATH

Department of Economics

Faculty of Humanities and Social Sciences

Dissertation for the Degree of MSc Economics and Finance

US Inflation Forecasting: An Empirical Study

Marina Solomou

September 2020

Faculty of Humanities and Social Sciences

Any student found to have cheated or plagiarised in assessment will be penalised. The Board of Examiners for Programmes will determine the nature and severity of the penalty but this may mean failure of the unit concerned or a part of the degree, with no provision for reassessment or retrieval of that failure. Proven cases of plagiarism or cheating can also lead to disciplinary proceedings as indicated in University Regulation 7.

I am aware of the guidelines on plagiarism: this coursework is the product of my own work.

Name: ………Marina Solomou…………………………………………………………..

Signed: ……MSolomou………………………………………………………………….

Degree: ………MSc Economics and Finance…………………………………………….

Supervisor: ……Dr. Tao Huang…………………………………………………………

Word Count: …………14,934……….. **Date**: 24/09/2020 ………

This dissertation may be available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Copyright of this dissertation rests with the author. No quotation from the dissertation and no information derived from it may be published without the prior written permission of the author.

Abstract

The main objective of the current paper focuses on US inflation forecasting using a vast range of state-of-the-art mode specifications with different properties. Specifically all examined model specifications collapse to a Time Varying Parameter regression regime ensuring the meaningfulness of the results. The data sample employed composes of 129 real-activity macroeconomic indicators incorporated as explanatory variables in the regression specifications ranging from $1960M1 - 2017M12$. Additionally, the inflation measurements examined include the Consumer Price Index (CPI) and Personal Consumption Expenditures Price Index (PCEPI). Two empirical applications are conducted, the former focuses on the evaluation of monthly inflation forecasts while the latter incorporates a quarterly forecasting regime. Furthermore, for the informative performance evaluation two performance metrics are estimated, namely the Mean Squared Forecast Error (MSFEs) assessing the performance of point forecasts and the Average Logarithmic Predictive Likelihood (ALPL) assessing the predictive accuracy of the out-of-sample rolling window forecasts. Both measurements are expressed in comparison to a univariate AR(2) benchmark model. Lastly, the empirical findings of this paper signify the superior forecasting performance of high-dimensional model specifications incorporating dynamic variable and model selection strategies that enable the continuous assessment of exogenous predictors relevance in each simulation period, namely - TVP-BMA, TVP-GAMP and VBDVS models.

Keywords— Bayesian shrinkage, Stochastic volatility, high-dimensional interference, Belief Propagation; factor graph, time-varying parameters, dynamic liner modeling, dynamic variable selection, approximate posterior interference, forecasting

Table of Contents

List of Figures

List of Tables

List of Abbreviations

US .... United States

CPI .... Consumer Price Inflation

PCEPI .... Personal Consumption Expenditures Price Index

VAR .... Vector Autoregressive

ARMA .... Autoregressive Moving Average

DSGE .... Dynamic Stochastic General Equilibrium

MCMC .... Markov Chain Monte Carlo

FGR .... Factor Graph Representation

GAMP .... Generalised Approximate Message Passing

TVP .... Time Varying Parameter

CLT .... Central Limit Theorem

SBL .... Sparse Bayesian Learning

iid .... Independent and identically distributed

AWGN .... Adaptive White Gaussian Noise

MFVB .... Mean Field Variational Bayes

BP .... Belief Propagation

DGP .... Data Generating Process

FRED .... Federal Reserve Economic Data

PCEPI .... Personal consumption expenditures price index

MSFE .... Mean square forecast errors

UC-SV .... Unobserved Component Stochastic Volatility

TVD .... Time Varying Dimension

SSVS .... Stochastic Search Variable Selection

DMA .... Dynamic Model Averaging

FFKF .... Forgetting Factor Kalman Filter

VB .... Variational Bayes

ELBO .... Evidence Lower Bound

DVS .... Dynamic Variable Selection

PIP .... Posterior Inclusion Probability

GPR .... Gaussian Process Regression

ELN .... Elastic Net Algorithm

PLS .... Parameter Least Squared

VBDVS .... Variational Bayes Dynamic Variable Selection

LASSO .... Least Absolute Shrinkage and Selection Operator

CSAPLD .... Cumulative Sums of Average Predictive Likelihood Differentials

1 | Introduction

Inflation control is centered at the very core of modern monetary policy making, thus a great attention has been attributed over the years by macroeconomists and central bankers in fabricating reliable inflation forecasting specifications for the achievement of policy targets. Additionally, the life evolution of the general price level over a commitment's lifetime is vital for the private sector's decision making processes hence, precise inflation forecasts are required for numerous reasons, namely - the evaluation of future policy implementations and the formation of economic agents inflation expectations for the optimal negotiation of long-term nominal commitments including, labor contracts, wages and mortgages.

However, according to ample empirical evidence, the price inflation of United States (US), depending on one's perspective has become both easier and harder to forecast. In light of this, various macroeconomic series, including inflation, appear to be considerably less volatile in the last decade compared to the period of 1970s and 1980s. Consequently the the root mean squared error associated with naive inflation forecasts exhibits a steep decline. In that respect, inflation forecasting can be thought as less cumbersome since the associated hazard, estimated via the mean squared forecast errors (MSFE), have declined significantly.

On the contrary, the evolution of standard multivariate forecasting specifications namely the backward-looking Phillips curve, in regards to a univariate benchmark model appeared to be less effective in capturing all inflation dynamics now, compared to the time period before 1980s. The underlying statement was forcefully emphasised by the study of Atkeson et al. (2001) whose findings endorsed the argument that since 1984 the US inflation forecast evaluations of the Backward looking Phillips curve, were classified as inferior to the forecast realisations associated with much more simplistic model specifications including the simple average rate of the former 12 months. Hence, in that sense the provision of further forecasting value over the univariate model became much more challenging and demanding. One can speculate on the economic rationales that stipulated this phenomenon but for a more informative conjecture, the precise reasons and modifications in inflation process that gave rise to the changing properties of inflation forecasting are required to be initially outlined.

First and foremost, changes in inflation process can be succinctly related to the variation of the Phillips curve slope Blanchard (2016) in which case, one can assume a similar variation in the predictors coefficients. Consequently, the determinant predictors of the slope estimations inevitably undergo chronic changes in terms of forecasting relevance. An enormous macroeconomic literature body provides evidence in regards to the structural instability characteristics of macroeconomic time series exploiting further the time variation among explanatory parameters. The adverse consequences of ignoring the underlying phenomenon for inference and forecasting has extensively been stressed out by inter alia Clements & Hendry (1998),Koop & Korobilis (2012) and Pesaran et al. (2006). Thus the necessity of incorporating such instabilities in forecasting model specifications has been extensively motivated (Stock Watson (1996)).

Secondly, the explanatory predictor assortment relevant to the examination objective can vary significantly. For instance, in his empirical study Groen et al. (2013) incorporates a total number of 10 macroeconomic predictors. However, economists working with factor modeling including Stock Watson (1996) often employ a significantly larger predictor set. Assuming a potential predictor sample of size m, the total statistical model number occurring equals to 2^m resulting into increasing substantially both the statistical and computational difficulties of model selection strategies which may potentially lead to model mispecifications. As a response to the underlying implication macroeconomists have turned to the Bayesian interference implementing either an automated model selection process or Bayesian model averaging (BMA) as encountered in the empirical work of Koop & Potter (2007). Furthermore, the underlying specification was originally implemented in the empirical work of McConnell & Perez-Quiros (2000) who entailed a singular change-point model specification providing adequate evidence in regards to the sharp fall of the volatility in the US activity during the early 1984. .

Thirdly, the relevance of a particular model specification may vary over time as well especially through different phases of economic cycles complicating further the already challenging macroeconomic exercise. For instance, assuming the implementation of a different forecasting model over different time periods Ï, this argument would potentially give rise to a total number of 2mÏ models combinations which is essentially computationally infeasible to forecast.

Nevertheless, present posterior simulation algorithms are in an inferior position to broaden the scope of high-dimensionality due to the significantly large numerical error associated and computational inefficiency arising through repetitive Monte Carlo simulation samples Angelino et al. (2016). In light of this, although Bayesian interference provides a natural probabilistic regime enabling the exploitation of all relative parameters via the entailment of all available information in the data prior and likelihood, real-time decision making in multiple-dimensions is unsupported by the arising computational limitations.

Thus, it comes as no surprise that the types of approaches and model specifications suitable for US inflation forecasting remain an open empirical question to the wider economic society. Hence, the main objective of the current empirical study is to shed light on this challenge via the examination and assessment of the performance of a spectrum of forecasting specifications for different time horizons. Precisely, the forecasted US inflation measures include the CPI and PCEPI via the implementation of 129 monthly real-activity explanatory variables ranging form $1960M1 - 2017M12$. The evaluation of the realised forecasting results and the models performance is based on two metrics. The former includes the conventional mean squared forecast error in relation to a benchmark AR(2) model and the latter is based on the complete range of predictive density known as average logarithmic predictive likelihood difference from the benchmark. Additionally, the empirical exercise is divided into two subsections, the first of which uses monthly data frequencies while the second focuses on quarterly inflation forecasting.

Following the introduction, Chapter 2 provides an overview of the key theories and concepts motivating the current investigation. Chapter 3 outlines the gathered data and their respective transformations and Chapter 4 sets forth some preliminaries and concepts of the investigated regimes. Furthermore, Chapter 5 outlines the many-predictors inflation forecasting exercise whose results are presented and discussed in Chapter 6 and finally, Chapter 7 concludes the key empirical findings.

2 | Literature Review

Inflation forecasting has been classified by the wider economic society as one of the most challenging yet important macroeconomic exercises. Numerous approaches and estimation regimes have been suggested with the most prevalent being based on extensions of the Phillips curve. Thus, the most influential and representative papers that impacted the literature of inflation forecasting in combination with the associated findings are outlined in the following discussion.

Employing the publication of Keynes's *General Theory* in 1936, followed by the disclosure of the reciprocal work of *Econometric model of the United States:1929-1925* by Klein & Goldberger (1955) and *Keynesian Revolution* 1955, as the general entry point of the current discussion, the period that followed was characterised by intellectual activity that pioneered the development, evaluation and analysis of Keynesian structural models. In addition, the statistical aspect of the associated research was stimulated by the innovative work of Neyman, Fisher, Pearson inter alia during that century. Following the Keynesian theory advances, structural Keynesian macroeconomic forecasting incorporating postulated regimes of decision making rules, relished a golden era during 1950s and 1960s. However, they did not succeed to live up to their original expectations as cracks in their foundations began to emerge during the great inflation and disinflation period, 1970s - 1980s deteriorating their forecasting performance.

As a response to the receding implementation of Keynesian structural macroeconomic forecasting, econometric regimes based on traditional system-of-equations were collectively augmented to rectify the theoretical defects at the time. A significant contribution that accelerated the technologically oriented statistical research was the Lucas' formal critique in regards to the well established system of equations approach Lucas et al. (1976), questioning the effectiveness of decision rules analysis in regards to the evaluation of conditional forecasts by emphasising that decision rule parameters are expected to alter with changes in policies. Furthermore, an alternative response that engaged a more radical change of framework was the introduction of nonstructural forecasting models. The fundamental basis of such models lies on the unconditional forecasting of the economic future pathway under the assumption of no structural changes in policy. Thus, in combination with the emerging Keynesian macroeconomic theory discontent, an immense interest grew in nonstructural forecasting regimes during the 1970s including the important work of Sargent et al. (1977).

However, the literature of nonstructural foresting regimes is traced back to late 1920s with the argument of Yule (1927) and Slutzky (1937) who signified the effectiveness of simple linear difference equations with the driving mechanism of randomised stochastic shocks, as a powerful tool for both the estimation and forecasting of numerous financial and macroeconomic time series. These stochastic equations are widely known as autoregressions and are interlinked with the classic moving average processes. The distinct feature of such process system dynamics includes the translation of randomised inputs into serially correlated outputs. The underlying phenomenon is known by the wider economic society as The Slutsky-Yule effect.

The mathematical H.Wold with his landmark contribution laid the foundation to the novel work of A.Kollogorov, N.Wiener and R.Kalman who demonstrated that under the assumption of adequate stability, the stochastic component of a time-series probabilistic mechanism generator can have a similar representation to the model specification form introduced by Slutsky and Yule. Consequently, autoregressive models were critical for optimal macroeconomic forecasting.

Additionally, a major contribution in the literature of nonstructural time series analysis includes the monument book publication *Time Series Analysis: Forecasting and Control* by Box et al. (1994). Various Box-Jenkins insights lead to the excitation of an explosively growing literature including the suggestion of cumulative random shock effects as the main trend driver, commonly known as "stochastic trend". The underlying idea and implication range of stochastic trends was further amplified by the empirical work of inter alia, Nelson & Plosser (1982) and Campbell & Mankiw (1987). Nevertheless, according to Diebold & Senhadji (1996), the long-run forecasts of the associated direct implementation, fail to converge to a specific trend thus, the redefinement of the trend location was necessary.

The centerpiece, of Box and Jenkins literature contribution however, lies in their fabrication and demonstration of an applied nonstructural forecasting operational framework comprising of a counter-cyclical regime of model formulation, evaluation, diagnostic testing and forecasting known as Autoregressive Moving Average (ARMA) models, enabling a more parsimonious dynamic approximation compared to both moving average and autoregressive models.

Innovative work emerged in the following period, extending upon Box and Jenkins contributions centered however around multivariate dynamic and cross-variable relationships. The econometric framework is based on the independent variable regression both on past values- lags of itself and other predictive variables in the system incorporating this ways cross-variable interlinks and controlling for correlations between regressors disturbances. An equally important multivariate contribution includes the notion of Granger-Sims causality tests originally introduced by Granger (1969) and Sims (1972).The underlying framework is applied for the examination of feedback mechanisms and crossed-linkage patterns in vector autoregression regimes. Furthermore, in a later publication of Sargent & Sims (1977) the concept of dynamic factor models was firstly introduced capturing the proposition of common and idiosyncratic economic shocks across various economic and financial sectors insinuating at the same time the dependence of large variable sets on distinct common sources of variation, common economic features and consequently common economic data. Thus, the parsimonious modeling and large predictor forecasting regimes were further facilitated by the comovements generated by common shocks. Key contributions of the underlying methodology include Stock & Watson (1989), Chan et al. (1999), Quah (1993) and Staiger et al. (1997).

The entire discussion up to this point was based on linear models. However, significant attention had also been attributed in nonlinear foresting regimes, an important example of which includes volatility dynamics models allowing for volatility forecasting. Seminal contributions in the related literature are namely the studies of Engle (1982), Bollerslev et al. (1994) and Bollerslev et al. (1994). The main stand of nonlinear literature lies in the concept of switching regimes for the approximation of business cycle stages. Thus more emphasis is attributed to tracing and chronologically constructing the underlying economic phenomenon for the identification and examination of turning points strongly associated with coincident indicators as illustrated by Diebold & Rudebusch (1994).

Regime-switching model specifications are classified as an innovative embodiment of the Burns & Mitchell (1946) non-linear forecasting regime. The fundamental idea of regime switching specifications is applied through various threshold models whereas a specified regime is determined by an explanatory variable. Specifically, the empirical applications of Granger et al. (1993) and Tong (1990) associate the indicative variable with an unobserved variable's historical aspect. Oppositely, Hamilton (1989) claimed that unobserved regime indicator models are superior for macroeconomic and financial forecasting. In particular, Hamilton's the innovative "Markov-switching" models engage an unobserved indicator for the regime determination process. A milestone extension in the literature of structural breaks was originally introduced by Chib (1998) who integrated the well-known Bayesian approach with structural break regimes via the application of a Markov-chain transition probabilities for the approximation of the breaking process.

In the years that followed, regime-switching forecast specifications flourished among the econometric society following the substantial change in time series characteristics of US inflation measures reported by the findings of Cogley & Sargent (2005). In light of this, significant time variations in both the inflation mean and persistence where signified by Cogley & Sargent (2001). Additional evidence emphasising the variance breaks experienced by macroeconomic variables emerged from the findings of Cogley et al. (2010) and Heij et al. (2004) who revealed that over the time span of $1959 - 1999$ the vast majority of US macroeconomic variables experience volatility reductions associated to breaks in conditional volatility rather than conditional mean. Complementarily, Sims & Zha (2006), integrated the concept of Markov-switching time variation into a structural Bayesian Vector autoregression regime and according to their empirical findings, the time-variation experienced by the US macroeconomic dynamics was attributed solely to variance shocks breaks rather than breaks on regression parameters. Alternatively, intercept shifts were identified as the main forecasting failure as per Clements & Hendry (1998) and the wide-spread nature of parameter instability in linear regression models was further denoted by Stock & Watson (1996).

The outlined macroeconomic series features gave rise to an innovative wave of structural break literature with the contribution of Pesaran et al. (2006) further extending the structural Bayesian forecasting approach via the introduction of hierarchical parameter priors. However, the forecasting performance of regime switching models with a small number of structural breaks experience some limitations. According to literature, without the presence of restrictions, the number of coefficients can change significantly following a structural break. Furthermore, the forecast generation process does not account for future structural breaks following the last break identified in the data sample. These limitations ultimately compromise the forecasting performance of such models giving rise to estimation discontinuities.

Hence, a significant advancement was further introduced by the innovative work of Koop & Potter (2007) addressing the aforementioned limitations of the original change-break models. Specifically, the model specification introduced generally known as *KP-AR*, was enhanced with further desirable features namely - $i)$ no *ex ante* restriction implementations on the regime number and maximum duration $ii)$ neither constant nor monotonic approximation in regards to the regime duration evolution, $iii)$ parameter interconnectedness between adjacent regimes, $iv)$ duration information exchange between past and future regimes. The examined specification was originally implemented in the forecasting exercise of US GDP growth and CPI inflation using quarterly observations ranging from $1974Q2 - 2005Q4$ entailing as explanatory variables the first two lags of the dependent variable and an intercept. Finally, according to the original paper's empirical findings, the KP-AR model outperforms the simplistic one-break model specification convincingly denoting the advances of employing a random break number in change-point specification.

An additional contribution to the Bayesian structural break literature tackling the interference and modeling problem of time-series processes exhibiting arbitrary parameter changes at arbitrary "time-points" was introduced by the empirical investigation of Giordani & Kohn (2008). Specifically, the first contribution of the paper includes the employment of a mixture innovation in state variables regime for the approximation of the structural break process including the desirable features of efficient interference performance for time-series undergoing breaks in conditional variance. Furthermore, the discrete latent variable models sampling regime of the adaptive Metropolis-Hasting algorithm was implemented improving sampling efficiency in various model specifications entailing change-point, regime-switching, mixture-innovation and outlier detection. The underlying *GK-AR* specification was originally used by the respective authors for the quarterly US CPI inflation forecasting using inflation observations from the period $1951Q1-2004Q4$. According to the empirical findings, the examined mixture innovation approach represents a natural and universal way of rigorous and efficient evaluation of out-of-sample point CPI point forecasts.

However, standard statistical analysis such as regression analysis, often proceeds conditionally on a specific statistical model chosen through a pool of several competing model specifications with the objective of forecasting certain time-series data. Nevertheless, accounting for the wide range of plausible generated estimations from alternative model specifications the drawn results are always characterised by a certain element of uncertainty. Thus for the elimination of such uncertainties Leamer & Leamer (1978) introduced for the first time the Bayesian model averaging framework. The fundamental concept of the underlying regime depends on the statistical analysis on the entire sample spectrum of available statistical modes specifications via the implementation of model averaging and combination tools. In particular, Wright (2009) employed the BMA regime for the forecasting application of US CPI and PCE deflator using 30 explanatory variables and averaging across 93 different model specifications. According to the author's empirical findings BMA quarterly out-of-sample inflation forecasts were superior to simple averaging approximations outperforming the same time the AR model specification.

An additional model combination approach was later introduced by Tibshirani (1996) for the purpose of improving the statistical predictability of point forecasts and interpretability of forecasting model specifications by modifying the model fitting process to identify the optimal subset of the realised covariances which is implemented on the finalised model specification rather than employing the entire set. Furthermore, the underlying regime substantially reduces the forecasting error via the shrinkage of significantly large predictor coefficients by forcing the total sum of their absolute value to be less than a threshold, eliminating any potential overfilling issues. In their empirical application of Brazilian CPI inflation forecasting Inoue & Kilian (2008) denoted the forecasting gains of the so called LASSO specification especially in regards to smaller evaluation horizons.

A further enhanced BMA forecasting specification incorporating both instability in the relationship between predictor variables and inflation measurements and an uncertainty element in regards to the inclusion probability of predictor variables in the regression specification was originally introduces by Groen et al. (2013) known as TVP-BMA. The main feature of the underlying model is that unlike the aforementioned Bayesian averaging model which solely engages model uncertainty, the examined regime averages over the entire regression spectrum incorporating every possible predictor variable combination, optimising this way the uncertainty element in regards to variable inclusion. Furthermore, additional instability features are allowed via the structural break approximation regime for both the error variance and the regression parameter terms incorporating the two key sources of inflation uncertainty as per Stock & Watson (1999). In the original paper, the authors evaluate the forecasting performance of the *TVP-BMA* specification through the forecasting application of US gross domestic product deflator (GDP) and personal consumption expenditure PCE deflator using quarterly data of the period $1960Q1 - 2011Q2$ and a subset of real-time predictors consisting of 15 macroeconomic time series. According to the author's empirical findings, the proposed model break specification successfully captured the time-varying inflation features reported by former studies adapting parsimony of each regression specification to the examined forecasting horizon. Finally, the highly accurate point forecasts of the empirical exercise indicated the the TVP-BMA model superiority in regards to the competing mode specifications.

An equally important contribution in the forecasting literature includes the study of Chan et al. (2012) where a novel dynamic mixture forecasting approach was originally introduced, dealing with the over-fitting issues of high-parameter regimes. Specifically, the examined methodology known as Time Varying Dimension (TVD) is based on the estimation of state-space models subject to states equality restrictions and is part of the growing literature that entails restriction implementations on the Time Varying Parameter (TVP) regression further explained in Chapter 4 in order to eliminate risks associated with overparameterization ensuing parsimony by allowing regression dimensions to vary over time. Once again, the forecasting performance of the underlying regime was evaluated by the respective authors in an empirical forecasting exercise involving US inflation using real-time forecasting data ranging from $1962Q1 - 2008Q3$. The empirical evidence of the aforementioned paper conclude the exceptional forecasting performance of the examined regime in comparison to constant coefficient models and popular structural break specifications displaying increasing forecasting gains for longer time horizons demonstrates both desirability and feasibility.

A different forecasting approach that incorporates variable uncertainty was further developed by Koop & Korobilis (2012) known as Dynamic Model Averaging (DMA). The key feature of the underlying forecasting framework lies in the fact that DMA models do not solely allow coefficient changes but also enable the complete forecasting model to vary over time shifting the marginal effects of the explanatory variables accordingly via the inclusion of forgetting factors. The forecasting performance of the underlying specification was evaluated via the empirical application of US inflation forecasting using a Generalized Phillips Curve specification and a data-set of 16 macroeconomic predictors spanning through data from $1960Q1-2004Q4$. Moreover, according to the authors original findings the examined regime displayed increased forecasting gains compared to competing model regimes enabling both abrupt and gradual changes in predictors explanatory performance.

Furthermore, motivated by the empirical findings of Atkeson et al. (2001) and Roberts (2004) signifying the flattening of the Phillips curve and the shift in the lagged inflation coefficients in the unemployment rate Phillips curve specifications that emerged during the mid 1980s Stock & Watson (2007) developed an alternative univariate model specification for the purpose of US inflation forecasting known as Unobserved Component Stochastic Volatility (UC-SV) model extending on the generalisation of unobserved component model specifications by approximating the time evolution of the logarithmic variances of both the permanent and transitionary disturbances as a Gaussian random walk. The forecasting performance evaluation of the UC-SV regime was accomplished via the empirical exercise of US inflation forecasting using quarterly averages of monthly inflation data spanning through $1960Q1 - 2004Q4$ and increased forecasting gains confirmed the model's suitability for inflation forecasting.

Moving forward, an innovative Bayesian specification for inference in dynamic regression models was originally proposed by Kalli & Griffin (2014) incorporating structural changes in both the coefficients and inclusion importance of the explanatory variables aiming the development of a more parsimonious TVP representation for macroeconomic forecasting. In particular the respective authors introduce a novel prior specification which enables the shrinkage of the regression coefficients to change accordingly with time along with a highly efficient Markov chain Monte Carlo approach for posterior inference. The novel specification known as Time Varying Sparsity (TVS) was applied in an empirical forecasting exercise of US inflation implementing 31 real-activity predictors ranging from $1965Q1 - 2011Q1$ and according to the empirical results the examined approach appeared to outperform competing Bayesian methods.

Lastly, a novel approach has been employed by the current investigation contingent on the econometric literature of factor graphs originally introduced by Kschischang et al. (2001). The underlying theory motivates the application of an easily maintainable and computationally efficient Bayesian evaluation algorithm. Whilst graphs often resemble a structure that portraits the general relationship of two variables, a Factor Graph Representation (FGR) achieves the portrayal of a multi-dimensional vector consisting of various model parameters, delineating the dynamics between all scalar factors with respect to their joint posterior distribution. Put differently, a FGR exemplifies a visual tool applied for the decomposition of a high-dimensional posterior distribution into smaller scale components and is often utilised for the development of both various *message passing algorithms* and multiple versions of parallel MCMC algorithms - the former of which is being emphasised as one of the key examination points of the current study.

More specifically, the adopted estimation strategy by the underlying investigation is generally known as the *sum-product algorithm* and is characterised by an exceptional computational efficiency compared to mainstream statistics Wand (2017). The fundamental idea of the examined concept can be summarized as a general rule in factor graphs that permits the estimation of marginal posterior distributions in an iterative manner. Thus, the application of this concept to a parametric system characterised by both arbitrary prior and likelihood functions results in a rather unambiguous solution. Thereby, the employment of a *Generalised Approximate Message Passing* (GAMP) algorithm originally proposed by Kamilov et al. (2011), is motivated for the inclusion of additional quadratic and Gaussian approximations of the derived solutions.

3 | Data

All data employed in the underlying study are extracted from the Federal Reserve Economic Data *FRED* website of St.Louis Federal Reserves Bank `https://fred.stlouisfed.org` . The original time span of the raw data ranges between $1960M1 - 2017M12$, thus using monthly observations the total sample size comprises of 696 observations for each macroeconomic variable in the data sample. Additionally, all examined variables are transformed to induce stationarity and all series are deseasonalised filtering out all possible stochastic and deterministic trends. The stationarity transformations implemented closely follow the established methodology on the basis of the relevant literature as introduced by Stock & Watson (1999). Specifically, assuming that $z_{i,t}$ denotes the initial unprocessed data series of variable z expressed in levels, when implemented in the algorithm as a predictor it undergoes the appropriate transformation indicated by the analogous transformation code *Tcode*, all of which are outlined in Table 3.1. However, accounting for the stationarity transformations along with the implementation of lag observations in the empirical application the effective sample decreases considerably. In regards to the second forecasting exercise, quarterly observations are estimated by the average value of three executive monthly observations for the purpose of further inflation forecasting using a different frequency sample.

Table 3.1: Series Transformation Description.

Tcode	**Transformation**	**Expression**
1	No transformation - Level	$x_{i,t} = z_{i,t}$
2	First Difference	$x_{i,t} = z_{i,t} - z_{i,t-1}$
3	Second Difference	$x_{i,t} = \Delta z_{i,t} - \Delta z_{i,t-1}$
4	Natural Logarithmic	$x_{i,t} = log z_{i,t}$
5	First difference of natural logarithmic	$x_{i,t} = log z_{i,t} - log z_{i,t-1}$
6	Second difference of natural logarithmic	$x_{i,t} = \Delta log z_{i,t} - \Delta log z_{i,t-1}$

According to literature, asset prices are forward looking economic variables. Thus, classified as highly representative predictors of future real economy and inflation as well as leading indicators for both contraction and expansion periods in the US economy. Specifically, there are five standard macroeconomic measures used for inflation forecasting namely - the CPI for all items (CPIAUCSL), the core CPI excluding energy and food (CPILFESL), the GDP price deflator (PGDP), the personal consumption expenditure price index (PCEPI) and the personal consumption expenditure deflator (PCE) (Stock & Watson (2008)). Hence, the main objective of the current empirical investigation focuses on forecasting both the total CPI and PCEPI macroeconomic series whose original observations are illustrated by Figure 3.1 and Figure 3.2 respectively. The rationale behind the choice of the underlying variables lies in the ability of the PCEPI measure to trace the price variations in the entire spectrum of goods and services purchased by all consumer economic agents capturing the prices average percentage change amongst the complete range of personal consumption expenditures unlike similar US inflation measures. Similarly, the CPI measurement captures price level changes of the average weighted consumer goods and services market basket purchased by economic households identifying changes in their spending patterns. Additionally, it is standard practice for both indices to be estimated in an either monthly or quarterly basis.

Furthermore, one of the additional focal points of the current investigation includes the examination of the relevant information enclosed in economic activity variables and their respective effectiveness in capturing the dynamics of inflation moments. In particular, the model specifications implemented in the following empirical applications incorporate a total number of 129 explanatory variables, to which the dynamics of inflation distributions are conditioned, covering a wide spectrum of all major sectors describing the aggregated real economic activity classified as follows - the Primary, Secondary, Tertiary and Quaternary sector as per Kenessey (1987).

The current examination however distinguishes all explanatory variables into the following subcategories: Exchange Rates and Interest Rates, Housing Starts and Permits, Income and Output, Labor Market, Money and Credit, Prices, Stock Market and finally Orders, Consumption and Inventories, as illustrated by the Appendix Table A.1. According to ample evidence, exchange rate financial indicators are a transmission channel through which inflation is potentially imported in open economies. Specifically, the empirical findings of Gordon (1997) and Gordon & Stock (1998) emphasise that exchange rates are statistically significant predictors of the conventional Phillips curve specifications. Similarly, Bernanke & Mishkin (1992) confirm the appropriateness of Federal Funds rate for short-term inflation forecasts estimations confirming their predictive content. Further informative parameters reflecting the investigated monetary policy state are also employed in the Money and Credit subcategory including the total reserves and M1/M2 money stock indices.

In addition, the housing sector also appears to constitute as a significant component of the total aggregate wealth imposing a substantial weight in various countries CPI measurements. Hence, measurements of real activity in the underlying sector have been classified as leading indicators for the US economic activity as suggested by the findings of Stock & Watson (1999). Moreover, employing the simplistic notion of price valuation being equivalent to the expected future earnings value, it is natural for one to assume the predictive content of market indexed in terms on inflation forecasting. The empirical interlinkage between economic activity and stock prices performance traces back to the empirical findings of Mitchell & Burns (1938).

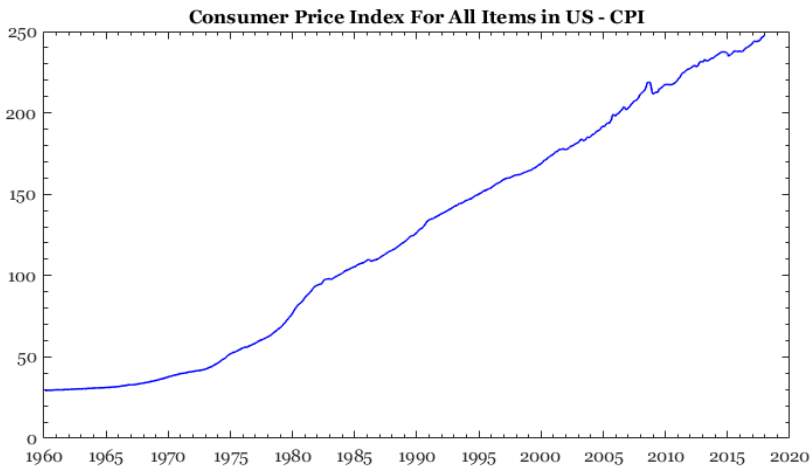


Figure 3.1: Seasonally Adjusted monthly observations of CPI ranging from $1960M1 - 2017M12$.

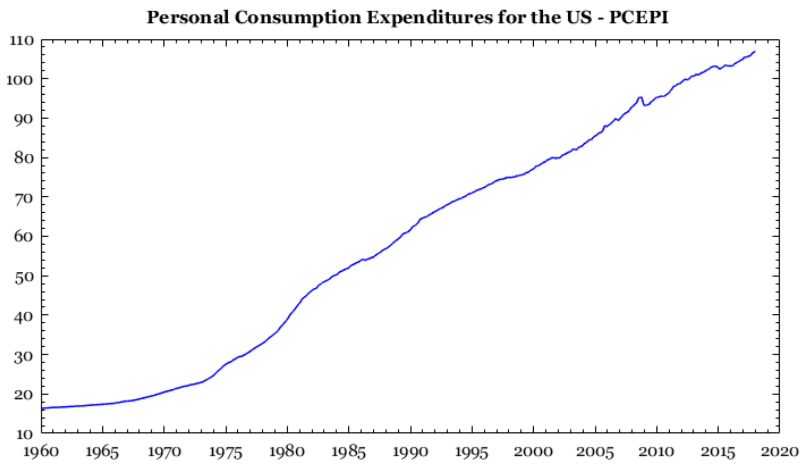


Figure 3.2: Seasonally Adjusted monthly observations of PCEPI ranging from $1960M1 - 2017M12$..

4 | Background

4.1 Sum Product Algorithm in a Factor Graph Regime

By definition, a factor graph represents the decomposition process of a global function consisting of various parameters into the product of smaller scale functions known as "factors". Employing the generic example where x is expressed as a function of discrete variables $x = (x_1, x_2, x_3)$, the decomposition of their joint mass function p is denoted by:

$$p(x_1, x_2, x_3) = \phi_a(x_1)\phi_b(x_1, x_2)\phi_c(x_2, x_3)\phi_d(x_3) \quad , \tag{4.1}$$

where $\phi_a, \phi_b, \phi_c, \phi_d$ signify the functions whose functional forms have are determined. As illustrated by Figure 4.1, factors are depicted by the filled boxes and random variables places are denoted by the circles. In addition, the estimation of the marginal distribution of variable x_i, requires the integration over all remaining variables x_i (summation in the discretised regime), such that:

$$p(x_i) = \sum_{\frac{x}{x_i}} p(x_1, x_2, x_3) \quad , \tag{4.2}$$

where $\frac{x}{x_i}$ signifies the subgroup x from which all x_i observations have been withdrawn. Assuming variable x consists of a total states number n, the respective summation operator adequately comprises of 2^{n+1} operations. Hence, in a high-dimensional regime with a substantial number of total variables/states computational effort arises significantly. Nevertheless, the disentanglement of variables is to some extent achievable, enabling the simplification of the summation operation and a reduced amount of algorithmic operations by substituting the expression of $p(x_1, x_2, x_3)$ in Equation (4.1) into Equation(4.2). For instance, according to the schematic representation variable x_1 appears to be indirectly linked to variable x_3,

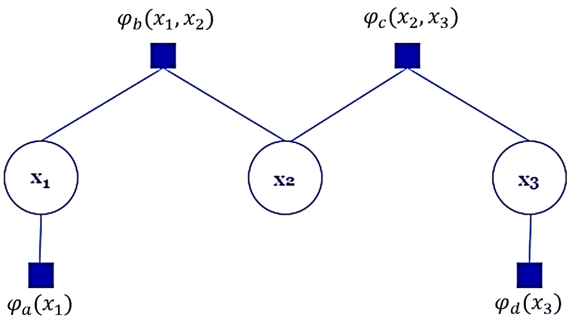


Figure 4.1: Simple factor graphical representation of joint function decomposition $p(x_1, x_2, x_3)$.

such that

$$p(x_1) = \sum_{x_2}\sum_{x_3} \phi_a(x_1)\phi_b(x_1, x_2)\phi_c(x_2, x_3)\phi_d(x_3) \quad , \tag{4.3}$$

$$= \phi_a(x_1)\sum_{x_2} \phi_b(x_1, x_2)\sum_{x_3} \phi_c(x_2, x_3)\phi_d(x_3) \ . \tag{4.4}$$

Hence, the employment reason of FGR is ultimately the complete representation of all potential influence channels that each individual system variable x_i exercises on the remaining system variables. Subsequently, the synopsis process of all possible pathways allows the optimal selection of all relevant factors ϕ_j for the estimation of marginal distributions enhancing this way the model's computational efficiency, which classifies the *message passing algorithms* as a suitable tool for marginal evaluations. More explicitly, the message propagating through each variable's node to the following interconnected variable can be formalised as a real-valued function encoding the influence of the underlying variable i the system. Therefore, in a multivariable high-dimensional regime the underlying methodology is superior to sequentially iterated conventional algorithms as it allows the estimation of parallel posteriors, thereby is frequently encountered in Bayesian machine learning applications Gelman et al. (2014).

4.1.1 Message Passing Interference

Indicating the propagating message from variable node x_i to function node ϕ_j as $\mu_{x_i} \rightarrow \phi_j$ and vice-versa $\mu_{\phi_j} \rightarrow x_i$, where $i = 1, 2, 3$ and $j = a, b, c, d$, the message propagating from x_i to ϕ_j is given by the superposition of all messages delivered to node x_i from which the message propagating from the target node ϕ_j is withdrawn, expressed by:

$$\mu_{x_i} \rightarrow \phi_j = \prod_{k \epsilon N(x_i), k \neq j} \mu_{\phi_k} \rightarrow x_i \ , \tag{4.5}$$

where $N(x_i)$ denotes the neighbouring nodes set of x_i. Accordingly, the message travelling from factor node ϕ_j to x_i is expressed as the product sum of all messages being received at that node and the factor function ϕ_j from which the message from the target variable node is excluded, signified by:

$$\mu_{\phi_j} \rightarrow x_i = \prod_{l \epsilon N(x_i), l \neq i} \mu_{x_l} \rightarrow \phi_j \ . \tag{4.6}$$

The algorithm class that encodes the continuous iteration between equations of similar form to Equations (4.5) - (4.6) is known as *sum-product algorithms* Pearl (1982). Belief propagation equations determine the iteration characteristics of the sum-product algorithm used for the estimation of marginal distributions. Consequently, the marginal distribution of variable x_i upon convergence is expressed as the product of all propagated messages from interconnected factor nodes expressed as:

$$p(x_i) \ = \prod_{m \epsilon N(x_i)} \mu \, \phi_m \rightarrow x_i \ . \tag{4.7}$$

Additionally, the general condition describing x_i variables, also known as *external nodes*, such as in the case of x_1 and x_3, is $\mu_{x_i} \rightarrow \phi_j = 1$. Similarly, ϕ_j is an external factor node and the general condition that holds for such cases is expressed by $\mu_{\phi_j} \rightarrow \phi_j(x_i)$.

4.2 Time-Varying Parameter Regression

The key reference point of the current section includes a stochastic volatility TVP regression expressed as:

$$y_t = x_t\beta_t + \varepsilon_t \ , \tag{4.8}$$

where y_t signifies the t^{th} observation value of the examined variable and t denotes all discretised time steps $t = 1, 2, 3, ...T$ with T being the aggregated examination time frame. The initial boundary condition implemented at $t = 0$ with regards to the $p \times 1$ vector of coefficients β is expressed as $\beta|_{t=0} = 0 = \beta_0$. Accordingly, x_t indicates the predictors' vector p with dimensions $1 \times p$, and $\varepsilon \sim N(0, \sigma_t^2)$, where σ_t^2 represents the stochastic time-varying variance parameter. Rather than making an assumption regarding the initial condition of the examined model, the corresponding non-centered parametrisation model has been implemented instead, dividing all system parameters into two subsections the former of which is assumed to be invariant capturing the initialised boundary condition and the latter of which operates as a supplementary time-varying component with 0 initial value for the avoidance of any discontinuities. Hence, the analogous regression specification is denoted by:

$$y_t = x_t\tilde{\beta} + x_t\tilde{\beta}_t + \varepsilon_t \ , \tag{4.9}$$

where the following boundary conditions $\tilde{\beta}_t|_{t=0} = 0$ and $\beta = \tilde{\beta} + \tilde{\beta}_t$ hold through the large-scale forecasting application. Specifically, the outlined parametrisation approach identifies whether the variables included in the system have either invariant coefficients or impose no explanatory power for estimating the dependent variable y, achieved by shrinking the time-varying component of the coefficient vector.

Therefore, employing a static representation, the TVP regression is denoted in an equivalent form as:

$$y = \mathscr{X}\beta + \varepsilon \ , \tag{4.10}$$

where $y = [y_1, y_2, ...y_T]'$ and $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_T]'$ are column vectors including all observations of y_t and ε_t respectively. Analogously, $\beta = [\tilde{\beta}', \tilde{\beta}_1', \tilde{\beta}_2', ...\tilde{\beta}_T']'$ is a $(T+1)p \times 1$ vector such that:

$$\mathscr{X} = \begin{bmatrix} x_1 & x_1 & 0_{1\times p} & ... & 0_{1\times p} & 0_{1\times p} \\ x_2 & 0_{1\times p} & x_2 & ... & 0_{1\times p} & {}_{1\times p} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ x_{T-1} & 0_{1\times p} & 0_{1\times p} & \cdots & x_{T-1} & 0_{1\times p} \\ x_T & 0_{1\times p} & 0_{1\times p} & \cdots & 0_{1\times p} & x_T \end{bmatrix}, \tag{4.11}$$

denoting a $T \times (T+1)p$ dimensional matrix. Evidence suggest that the entry values of the first p columns resemble a static regression with constant parameters whilst further columns introduce "time-dummies" in the regression system. Furthermore, accounting for the rank of the Gram matrix ($\mathscr{X}'\mathscr{X}$) of order T and the total amount of regression coefficients q given by $q = (T+1)p$, OLS regressions fails to estimate the values of all regression coefficients as degenerate solutions exist and regularisation appears to be indispensable for the estimation process. Consequently, to overcome such limitations a classic approach from engineering is adopted assuming that both the trajectories of β_t and $\tilde{\beta}_t$ undergo a random walk expressed as $\beta_t = \beta_{t-1} + \eta_t$ where $\eta_t \sim N(0, Q)$ for a symmetric $p \times p$ covariance matrix Q characterised as positively-defined. Hence, implementing the random walk regression enables the expression of the entire TVP regression in a conventional state-space regime allowing the extraction of all complementary information necessary for the calculation of β_t using the data available from both $\mathscr{X}$ and y. According to Bayesian statistical theory, the additional information extracted resembles a conditional hierarchical prior denoted as $p(\beta|\beta_{t-1}) \sim N(\beta_{t-1}, Q)$, that establishes an optimal level of shrinkage, where the estimation approach is supported by a simulation smoother normally of MCMC means as per Primiceri (2005).

The following section outlines the implementation of the shrinkage TVP regression model, adopting a reciprocating interference methodology originally developed by Tipping (2001). According to the underlying approach, interference is obtained irrespective to the idiosyncratic conditional hierarchical prior of β_t subject to β_{t-1} but rather the direct recovery of the TVPs is motivated based on Equation(4.10) in combination of observational hierarchical shrinkage priors. Specifically, each entry element β_i of vector β is assigned to an independent unique hierarchical prior such that:

$$p(\beta_i|\alpha_i) = N(0, \alpha_i^{-1}) \tag{4.12}$$

$$p(\alpha_i) = Gamma(\underline{\alpha}, \underline{\beta}) \quad , \tag{4.13}$$

where Equation(4.12) signifies the conditional Normal prior regarding β_i and Equation (4.13) denotes the Gamma prior regarding α_i known as the precision parameter both displaying an alternative scale-combination representation of a Student's-t prior with exceptional shrinkage properties according to Korobilis (2013). Following the framework ofTipping (2001), for the replication purpose of the empirical results, the logarithmic notation of the uniform hyperpriors is adopted such that $\underline{\mathrm{a}} = \underline{\mathrm{b}} = 1$.

4.3 Bayesian Regression Representation in a Factor Graph Regime

This section expands on the factor graph form representation of the static TVP regression model. Employing the Bayes theorem the arising posterior expression for an independent β prior expressed as $p(\beta) = \prod_{i=1}^{q} p(\beta_i)$ is denoted by:

$$
\begin{aligned}
p(\beta|y) &\propto p(y|\beta)p(\beta) \ , \\
&= \prod_{t=1}^{T} p(y_t|\beta) \prod_{i=1}^{q} p(\beta_i) \ .
\end{aligned}
\tag{4.14}
$$

In addition, the marginal posterior of β_i with $1 \leq i \leq q$ is given in the following form:

$$
\begin{aligned}
p(\beta_i|y) &= \int p(y|\beta) \, d\beta_{j \neq i} \ , \\
&\propto \int p(y|\beta) \, p(\beta) \, d\beta_{j \neq i} \ , \\
&= p(\beta_i) \int p(y|\beta) \prod_{j \neq 1, j \neq i}^{q} p(\beta_j) \, d\beta_{j \neq i} \ ,
\end{aligned}
\tag{4.15}
$$

where the marginal posterior of β_i is derived by the integration over the entire range of the remaining $q - 1$ vector parameters signified as $\beta_j|_{j \neq i}$. Put differently, the aforementioned estimation stipulates the execution of a $(q - 1)$ - dimensional integral, thus in estimation models with a high-dimensional β vector the numerical simulation appears to be computationally trivial if not infeasible.

Recalling the factor graph framework the marginal posterior of β is effectively factorised via the application approach illustrated by Figure4.2 where the marginal posterior of β_i is expressed as the convolution of all messages propagating towards node β_i such that:

$$
p(\beta_i|y) = \mu_{p(\beta_i) \to \beta_i} \prod_{t=1}^{T} \mu_{p(y_t|\beta) \to \beta_i} \ .
\tag{4.16}
$$

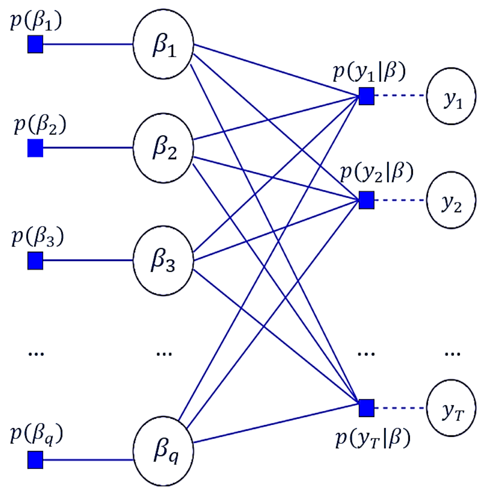


Figure 4.2: Illustration of high-dimensional regression model as a factor graph diagram.

Similarly to Figure 4.1, the $\mu_{p(\beta_i)\to\beta_i}$ message is defined as an external factor node thereby its value is no different to the prior $p(\beta_i)$. As a matter of principle, according to the sum-product rule, recalling Equations (4.5) - (4.6), the message propagating from $p(y_t|\beta)\forall t$ towards β_i is denoted by:

$$\mu_{p(y_t|\beta)\to\beta_i} = \int p(y_t|\beta) \prod_{j=1,j\neq i}^{p} \mu_{\beta_j\to p(y_t|\beta)}\, d\beta_{j\neq i} \;, \tag{4.17}$$

where $\mu_{\beta_j\to p(y_t|\beta)}$ represents the message propagating from β_j node to the factor function $p(y_t|\beta)$, defined as the product of the entire set of incoming messages from which the message propagating from $p(y_t|\beta)$ is excluded.

$$\mu_{\beta_j\to p(y_t|\beta)} = p(\beta_j) \prod_{s=1,s\neq t}^{T} \mu_{p(y_s|\beta)\to\beta_j} \;. \tag{4.18}$$

Due to the counter-cyclical interaction between Equation(4.17) and Equation(4.18), as one can observe that the message decomposition process of $\mu_{\beta_j\to p(y_t|\beta)}$, requires‘ the calculation of $\mu_{p(y_t|\beta)\to\beta_i}$ and vice-versa. Hence, both equations are being updated in a repeatedly expressed by the following iterative sun-product regime:

$$\mu^{(r+1)}_{p(y_t|\beta)\to\beta_i} = \int p(y_t|\beta) \prod_{j=1,j\neq i}^{p} \mu^{(r)}_{\beta_j\to p(y_t|\beta)} d\beta_{j\neq i} \ , \tag{4.19}$$

$$\mu^{(r+1)}_{\beta_j\to p(y_t|\beta)} = p(\beta_j) \prod_{\int=1,\int\neq t}^{T} \mu^{(r)}_{p(y_\int|\beta)\to\beta_j} \ , \tag{4.20}$$

where the algorithm's r^{th} iteration is signified by superscript (r). In the classic tree diagram structure the precise marginal posteriors of the system's β_i parameters are being retrieved after each algorithmic iteration. This however, might not necessarily be the case in terms of the sum-product method as convergence to an accurate value is not always ensured. However, according to Mooij & Kappen (2007) the examined sum-product algorithm is widely used in a diverse range of fields namely computational optimisation, computer vision and error correcting codes accomplishing reasonable approximations. Furthermore, the condition required for the establishment of the successful algorithmic convergence of Equation (4.10) includes a reasonably low correlation regarding the predictors of the right hand side. In the case where the condition fails to be met, the joint posterior of β coefficients experience high correlation resulting in an inaccurate interference solely depending on the marginal posteriors.

The incurred correlation of the benchmark TVP regression ranges by construction within reasonable limits as a result of the block-diagonal structure of the Gram matrix $(\mathscr{X}'\mathscr{X})$ that compensates for potentially inadequate correlation arrangements including high block correlations. Additionally, the predictor variables entailed in the following empirical application include either lags or principal components hence, it is uncommon for exceptionally high correlations to occur within blocks. Lastly, general time-series and autoregressive models are not being accommodated by the particular likelihood function time-decomposition, and empirical findings suggest the accurate recovery of the autoregressive coefficients.

4.3.1 Generalised Approximate Message Passing Model - GAMP

Following the thorough explanation regarding the core idea of message passing algorithms, the exact derivation of the functional form describing the propagating messages denoted by Equations (4.19) - (4.20) subject to the Student's-t hierarchical prior and the regression likelihood is a rather trivial task as it requires a cumbersome integration. To be precise, the examined algorithm incorporates a class of Gaussian approximations in to sum-product algorithm iterations premised on asymptotic results that enhance the reliability of the GAMP approximation especially in a high predictor number regime. On the fundamental basis of the Central Limit Theorem (CLT), in the case where $q \to \infty$ the message $\prod_{j=1mj\neq i}^{q} \mu_{\beta_j \to p(y_s|\beta)}$ is estimated via a Gaussian distribution approximation with reference to the uniform norm initially introduced by Berry (1941). Thereby, the message described by Equation(4.20) can be assumed to be analogous to a Gaussian distribution. Furthermore, the message terms are estimated via the Taylor series expansion capturing all required information for the analytical estimation of the first two moments of $p(\beta_i|y)$ with omission error magnitude being $O(\frac{1}{q})$. Recalling the findings of Rangan (2011) in the limits of $q \to \infty$ and $\frac{q}{T} \to \delta$ both approximation assumptions ease off. The underlying phenomenon is widely known as "the Big-Data Blessing' and is identified in a wide range of estimation algorithms facilitating a substantial number of q asymptotics such as GAMP.

Implementing all approximations in the two core algorithm update Equation (4.19) and Equation (4.20) results in the formulation of a finalised intelligible algorithm that estimates the approximated values of the first two moments of $p(\beta_i|y)$. The simulation process for the approximation of the marginal posterior distribution includes the cyclical iteration among computationally straightforward addition and multiplication operations giving rise to the algorithmic complexity of magnitude

$\propto O(Tq)$, resulting in an increasing efficiency of the algorithm regarding handling regressions characterised by an extremely large predictor number. The key condition required for algorithmic convergence to be reached is described by the difference between two sequential iterations posterior mean approximations of β to be less than a threshold value. Additionally, parametric updates including the value of hyperparameter a_i in Equation (4.12), are calculated via the combination of both the GAMP algorithm and EM updates (for more details please refer to Al-Shoukairi et al. (2017)). Finally, a simplified version of the algorithm employed is illustrated by *Algorithm 1*, outlining the estimating process of β under the assumptions of fixed normalised value assignments for both prior hyperparameters and regression variance.

Algorithm 1 Generalised Approximate Message Passing (GAMP) algorithm with estimated prior hyperparameters and variance

Initialization $\hat{\beta}_i^{(0)} = 0$, $\hat{\tau}_i^{\beta,(0)} = 100 \; \forall \, i = 1, ...q$ and
$\hat{s}_t^{(0)} = 0 \, \forall \, t = 1, ..., T.$
$r = 1$
while $||\hat{\beta}^{(r)} - \hat{\beta}^{(r-1)}|| \to 0$ **do**
1) *Output message step:*
FOR $t = 1$ **to** T **do**

$$\begin{aligned}
\hat{c}_t^{(r)} &= \textstyle\sum_{i=1}^{q} \mathscr{X}_{t,i}\hat{\beta}^{(r-1)} - \hat{s}_t^{(r-1)}\hat{\tau}_t^{c,(r)} \\
\hat{\tau}_t^{c,(r)} &= \textstyle\sum_{i=1}^{q} \mathscr{X}_{t,i}^2 \hat{\tau}^{\beta,(r-1)} \\
\hat{s}_t^{r} &= g_{out}(\hat{c}_t^{(r)}, \tau_t^{c,(\hat{r})}, y_t) \\
\hat{\tau}_t^{s,(r)} &= -\tfrac{\partial}{\partial \hat{c}} g_{out}(\hat{c}_t^{(r)}, \tau_t^{c,(\hat{r})}, y_t)
\end{aligned}$$

end for
2) *Input message step:*
FOR $i = 1$ **to** q **do**

$$\begin{aligned}
\hat{d}_i^{(r)} &= \hat{\beta}_i^{(r-1)} + \hat{\tau}_i^{d,(r)} \textstyle\sum_{t=1}^{T} \mathscr{X}_{t,i}\hat{s}_t^{(r)} \\
\hat{\tau}_i^{d,(r)} &= (\textstyle\sum_{t=1}^{T} \mathscr{X}_{t,i}^2 \hat{\tau}_t^{s,(r)})^{-1} \\
\hat{\beta}_i^{(r)} &= g_{in}(\hat{d}_i^{(r)}, \hat{\tau}_i^{d,(r)}) \\
\hat{\tau}_t^{\beta,(r)} &= \hat{\tau}_i^{d,(r)} \tfrac{\partial}{\partial \hat{d}} g_{in}(\hat{d}_i^{(r)}, \hat{\tau}_i^{d,(r)})
\end{aligned}$$

end for
$r = r + 1$
end while
Assign estimated values for mean and variance of β subject to:
$\hat{\beta} = (\hat{\beta}_1^{(r)}, ..., \hat{\beta}_q^{(r)})$ and $\hat{\tau}^{\beta} = (\hat{\tau}_1^{\beta,(r)}, ..., \hat{\tau}_q^{\beta,(r)})$

Having examined all necessary components, the algorithm consists of two main steps, the former includes the evaluation of the output signal comprising of all messages propagating from every variable node β_j and the latter includes the evaluation of the input signal composed by the complete message set of signals propagating towards each variable node β_j. The convoluted signal denoting the result function from both the input and output signals is used for the estimation of the first two moments, namely mean $\hat{\beta}_i$ and variance $\hat{\tau}_i$ of β_j subject to the two scalar nonlinear functions g_{in} and g_{out} whose analytical derivation is outlined in Appendix (B). Finally, it is worth mentioning that regardless of the functions' form the greatest complexity level of the examined algorithm is of magnitude $\propto O(Tq)$.

At this point the Bayesian stochastic volatility estimator originally introduced by Kim et al. (1998), is implemented as a variance estimator in the GAMP algorithm. This is done by expressing the main regression model in the following form:

$$y = \mathscr{X}\beta + \Sigma\nu \ , \tag{4.21}$$

where $\mathscr{X}$ indicates a $T \times T$ dimension matrix with main diagonal elements representing the time-varying standard deviations σ_t. Thereby, assuming estimates of β are obtained trough $\hat{\beta}$ approximations, the key model is expressed by the following logarithmic form:

$$log[(y - \mathscr{X}\hat{\beta})^2] = log(diag(\Sigma)^2) + log(\nu^2) \ , \tag{4.22}$$

$$\Longrightarrow \tilde{y} = \tilde{\sigma}^2 + \tilde{\nu} \ , \tag{4.23}$$

where $\tilde{\bullet}$ represents variables in a log-square form and $diag(\Sigma)^2$ denotes a vector of all σ_t^2 elements of size $T \times 1$. In addition, the intrinsic distribution of $\tilde{\nu}$ appears to follow a single freedom degree $log-\chi^2$ distribution, which as per Kim et al. (1998) can successfully be evaluated via the combination of 7 equally weighted Normal distribution with mean, variance and weight values of μ_i, V_i and π_i respectively, such that $\sum_i \pi_i = 1 \forall i = 1, 2.., 7$. Substituting the above regime, Equation(4.23) is expressed as:

$$\tilde{y} = \tilde{\sigma}^2 + u_i \quad , \quad i = 1, 2, ..., 7 \quad , \tag{4.24}$$

where $u_i \sim N(\mu_i, V_i)$. The estimator entailed for the description of the log-volatilities vector of magnitude $T \times 1$ is expressed as $E_i(\tilde{\sigma}) = \tilde{y} - \mu_i$. Hence, the finalised expression regarding the approximation at given time t is given by:

$$\hat{\sigma}_t^2 = exp\left(\sum_{i=1}^{7} \pi_i \frac{(\tilde{y}_t - \mu_i)}{7}\right) . \tag{4.25}$$

Recalling the results of Korobilis (2019), the obtained volatility estimator is comparable with the analogous volatility estimator of Kim et al. (1998). At last, it is worth mentioning that the employed algorithm not only establishes a sensible balance between evaluation accuracy and computational effectiveness but it is proven to be highly efficient in macroeconomic forecasting applications.

4.4 State-Space Models and Variational Bayes Interference

The current section expands on the Variational Bayes (VB) estimation methodology focusing on the evaluation of intractable posterior distributions by applying the relevant formulas and generic framework to the examined TVP regression specification. For further details of the examined concept one can refer to the following studies Beal (2003), Wang et al. (2016) and the monograph of Šmídl & Quinn (2006).

4.4.1 Fundamentals of Variational Bayes

Employing a data set y, along with latent variables s and θ where the former signifies all unobserved state variables namely time-varying error variances and time-varying explanatory variable coefficients and the latter denotes all remaining parameters including the state equation error covariances, the estimations of interest include the joint posterior denoted by $p(s, \theta|y)$ with a corresponding marginal likelihood $p(y)$ and associated data-parameter joint density distribution $p(y, s, \theta)$.

Nevertheless, due to the high degree of complexity associated with the distribution evaluation process, the standard approach of establishing an approximating density distribution $q(s, \theta|y)$ from a more simplistic distribution band $\mathcal{B}$ evaluated over the corresponding space range of parameters s and θ is applied. Hence, the primary objective of Bayes interference includes the estimation of the relevant posterior distribution $q(s, \theta|y)$ minimising the magnitude difference - "distance" from the actual $p(s, |y)$ distribution quantified by a distance metric known the Kullback-Leibler divergence Hershey & Olsen (2007) expressed by:

$$KL(q||p) = \int q(s, \theta|y) \log \left\{ \frac{q(s, \theta|y)}{p(s, \theta|y)} \right\} dsd\theta \tag{4.26}$$

Thus, the optimisation problem is denoted as follows:

$$q^*(s, \theta|y) = arg \min_{q(s,\theta|y) \in \mathcal{B}} KL(q||p) \ , \tag{4.27}$$

where q^* denotes the optimal q value. Following a straight-forward readjustment of the marginal likelihood logarithm one can obtain:

$$log\, p(y) = log\, p(y) \int p(s, \theta|y) dsd\theta = \int p(s, \theta|y) log\, p(y) dsd\theta \tag{4.28}$$

$$= \int p(s, \theta|y) \log \left\{ \frac{p(y, s, \theta/q(s, \theta|y)}{p(s, \theta|y)/q(s, \theta|y)} \right\} dsd\theta \tag{4.29}$$

$$= \int p(s, \theta|y) \log \left\{ \frac{p(y, s, \theta)}{q(s, \theta|y)} \right\} dsd\theta + KL(q||p) \ . \tag{4.30}$$

According to the above expression the function $KL(q||p) \geq 0$, the lower bound of the marginal likelihood is identified by the following argument:

$$\mathcal{G}(q(s, \theta|y)) = exp \left[\int q(s, \theta|y) \log \left\{ \frac{p(y, s, \theta)}{q(s, \theta|y)} dsd\theta \right\} \right] \tag{4.31}$$

$$\equiv exp \left[\mathbb{E}_{q(s,\theta|y)} (\log(p(y, s, \theta)) - \log(q(s, \theta|y))) \right] \ , \tag{4.32}$$

where the $\mathcal{G}(q(s,\theta|y))$ is generally know as the Evidence Lower Bound (ELBO). Hence, the optimisation objective is modified to the identification of the approximating density $q^*(s,\theta|y)$ that maximises the ELBO and consequently the marginal density of the data $p(y)$. For the sake of simplicity the mean-filed factorisation identity described by $q(s,\theta|y) = q(\theta|y)q(s|y)$ is implemented transforming the optimal specification of both $q(s|y)$ and $q(\theta|y)$ to:

$$q(s|y) \propto exp\left[\int q(\theta|y)\log p(s|y,\theta)d\theta\right] \equiv exp[\mathbb{E}_{q(\theta|y)}(\log\ p(s,y|\theta)) \quad (4.33)$$

$$q(\theta|y) \propto exp\left[\int q(s|y)\log p(\theta|y,s)ds\right] \equiv exp[\mathbb{E}_{q(s|y)}(\log\ p(\theta,y|s))\ , \quad (4.34)$$

where the former equation represents the expectation of s conditional posterior and the latter signifies the expectation of θ's conditional posterior over $p(\theta|y)$ and $p(s|y)$ respectively. Due to the interconnectedness between $q(s|y)$ and $q(\theta|y)$, VB algorithms employ a counter-cyclical iteration regime between the two expressions until the optimisation condition is met maximising $\mathcal{G}$.

4.4.2 Dynamic Variable Selection -DVS

The key concept of the examined modelling regime includes the dynamic model selection strategy. Expanding on the "static" prior of variable selection introduced by George & McCulloch (1993), the specification of a static parameter regression via the application of MCMC is expressed as:

$$\beta_{j,t}|\gamma_{j,t},\tau^2_{j,t} \sim (1-\gamma_{j,t})N(0,\underline{c}\times\tau^2_{j,t}) + \gamma_{j,t}N(0,\tau^2_{j,t})\ , \quad (4.35)$$

$$\gamma_{j,t}|\pi_t \sim Bernoulli(\pi_{0,t})\ , \quad (4.36)$$

$$\frac{1}{\tau^2_{j,t}} \sim Gamma(g_0,h_0)\ , \quad (4.37)$$

$$\pi_{0,t} \sim Beta(1,1)\ , \quad (4.38)$$

where g_0, h_0 and $\underline{c}$ denote fixed valued hyperparameters. According to the fundamental variable selection principles, the initial component of the $\beta_{i,j}$ prior enforces the posterior shrinkage towards zero whilst the second component requires a sufficiently large variance $\tau_{j,t}^2$ for enabling unrestrained evaluation. Thus, for the underlying condition to be met $\underline{c}$ is required to be infinitesimally small $\underline{c} \rightarrow 0$. In addition, the component selection process in the $\beta_{j,t}$ prior relies on the random binary variable $\gamma_{j,t}$ characterised by a Bernoulli distribution so that in the case where $\gamma_{j,t} = 0$ the prior variance equals $\underline{c}\tau_{j,t}^2$ and in the case where $\gamma_{j,t} = 1$, $\beta_{j,t}$ is characterised by a normally distributed prior with zero mean and $\tau_{j,t}^2$ variance. As per the suggestions of Narisetty et al. (2014), the examined parameters are evaluated subject to the data dimensions T and p via different deterministic functions for the avoidance of any potential model inconsistencies. The current study follows the approach of Korobilis & Koop (2020), adopting the proposed boundary conditions including $\underline{c} = 10^{-4}$ so that the initial posterior component entails a small variance value and a random variable with Gamma prior representation of $(\tau_{j,t}^2)^{-1}$ ensuring the relevance of the updated parameter subject to all information enclosed in the data likelihood. In addition, the marginal prior of $\beta_{j,t}$ is approximated by a combination of Student's T distributions with excess kurtosis whose components enforce $\beta_{j,t}$ shrinkage irrespective to the $\gamma_{j,t}$ value. Overall, the outlined prior regime identifies various dynamic sparsity patterns imposing at the same time dynamic shrinkage in all time-varying explanatory variables. Finally, based on the outlined prior settings the mean probability of inclusion regarding all explanatory variables in the benchmark TVP specification at time t is denoted by $\pi = \mathbb{E}(p(\hat{\pi_{0,t}})) = 1/2$ and accordingly the associated posterior mean inclusion probability of predictor's j regression at time t, known as posterior inclusion probability (PIP) signified by $\tilde{\pi}_{j,t} = \mathbb{E}(p(\gamma_{j,t}|\mathbf{y}_{1:T}))$.

4.4.3 TVP Benchmark Specification

Employing the TVP benchmark model definition, and substituting Equation(4.35) into the state-equation, two conditional prior estimations of $\beta_{j,t}$ arise of the form:

$$\beta_{j,t}|\beta_{j,t-1}, w_{j,t} \sim N(\beta_{j,t-1}, w_{j,t}) \tag{4.39}$$

$$\beta_{j,t} + \gamma_{j,t}, \tau^2_{j,t} \sim N(0, v_{j,t}) \tag{4.40}$$

where $v_{j,t} = (1-\gamma_{j,t})^2 \underline{c} \times \tau^2_{j,t} + \gamma^2_{j,t}\tau^2_{j,t}$. Integrating the two prior expressions of β_t according to Wang et al. (2016) the state-space specification translates to:

$$\boldsymbol{\beta}_t = \tilde{\mathcal{B}}_t \boldsymbol{\beta}_{t-1} + \tilde{\boldsymbol{\eta}}_t \tag{4.41}$$

where $\tilde{\boldsymbol{W}}_t = (\mathbb{E}(\tilde{\boldsymbol{W}}_t)^{-1} + \mathbb{E}(\tilde{\boldsymbol{V}}_t)^{-1})^{-1} = diag(w_{1,t}, ...w_{p,t})$ such that $\tilde{\boldsymbol{\eta}} \sim N(0, \tilde{\boldsymbol{W}}_t)$, $\tilde{\boldsymbol{V}}_t = diag(v_{1,t}, ..., v_{p,t})$ and $\tilde{\boldsymbol{B}}_t = \tilde{\boldsymbol{W}}_t \times \mathbb{E}(\tilde{\boldsymbol{W}}_t)^{-1}$. The joint prior variance of $\beta_{j,t}$ under the employed regime is expressed as a function of parameters $w_{j,t}$ and $v_{j,t}$. Thus the finalised state-space representation of the dynamic variable selection TVP regression specification is expressed by:

$$y_t = x_t\beta_t + \varepsilon_t \ , \tag{4.42}$$

$$\beta_t = \mathcal{B}_t\beta_{t-1} + \tilde{\eta}_t \tag{4.43}$$

Due to the fact that the model specification solely relies on the unknown parameters $(\beta_{1:T}, w_{1:T})$ where the subscript $1:T$ denotes all the state variable observations ranging between $t = 1$ to $t = T$, the initialised independent prior specification employed is of the form:

$$p(\boldsymbol{\beta_0}, \boldsymbol{w_0}) = p(\boldsymbol{\beta_0}) \prod_{j=1}^{p} p(w_{j,0}) \tag{4.44}$$

$$= N(\boldsymbol{m}_0, \boldsymbol{P}_0) \times \prod_{j=1}^{p} [Gamma(c_{j,0}, d_{j,0})]^{-1} \ , \tag{4.45}$$

where $Gamma(a, b)$ indicates a Gamma distribution with corresponding shape and rate parameters a and b respectively. Furthermore, time prior t subject to all the data information is expressed via the Chapman-Kolmogorov Equation Karush (1961) such that:

$$p(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t-1}) = \int_{\mathscr{B}\mathscr{W}} p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}) \prod_{j=1}^{p} p(w_{j,t} | w_{j,t-1}) \tag{4.46}$$

$$\times\, p(\boldsymbol{\beta_{t-1}} \boldsymbol{w}_{t-1} | y_{t-1}) d\boldsymbol{\beta}_{t-1} dw_{1,t-1} ... dw_{p,t-1} \tag{4.47}$$

were $\mathscr{B}$ and $\mathscr{W}$ represent the support parameters of both $\boldsymbol{\beta}_t$ and $\boldsymbol{w}_t$ respectively. At last, the time posterior distribution is evaluated following the realisation of y_t measurement via the implementation of Bayes theorem as follows:

$$p(\boldsymbol{\beta}_t, \boldsymbol{w_t} | y_{1:t}) \propto p(y_t | \boldsymbol{\beta}_t, \boldsymbol{w_t}) p(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t-1}) \ . \tag{4.48}$$

Accounting for the analytical intractability of the outlined Bayesian joint posterior distribution, posterior conditionals are instead approximated via a Gibbs sampler, via the definition of an approximated variational density function imitating the examined time posterior t such that $p(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t}) \approx q(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t})$. Hence, the optimisation objective includes the identification of the specific hyperparameter set that accomplish the minimisation of the relative entropy function which, as previously discussed, is in line with the maximisation of the associated ELBO of the marginal likelihood logarithmic expressed by:

$$q^*(\beta_{t,t} | y_{1:t}) = arg \max_{q(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t})} \int q(\boldsymbol{\beta_t, w_t} | \boldsymbol{y}_{1:t} \log \left(\frac{q(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t})}{p(\boldsymbol{\beta}_t, \boldsymbol{w}_t | \boldsymbol{y}_{1:t})} \right) \tag{4.49}$$

Employing once again the mean field factorisation identity the sequential optimisation problem of both $\boldsymbol{\beta}_t$ and $\boldsymbol{w}_t$ is determined by the following counter-cyclical recursion iterations:

$$q(\boldsymbol{\beta}_t, |\boldsymbol{y}_{1:t}) \propto \exp\left(\int \log p(y_t, \boldsymbol{\beta}_t, \boldsymbol{w}_t|\boldsymbol{y}_{1:t-1}) \prod_j q(w_{j,t}|\boldsymbol{y}_{1:t}) d\boldsymbol{w}_t\right) \quad , \tag{4.50}$$

$$q(w_{j,t}, |\boldsymbol{y}_{1:t}) \propto \exp\left(\int \log p(y_t, \boldsymbol{\beta}_t, \boldsymbol{w}_t|\boldsymbol{y}_{1:t-1}) q(\boldsymbol{\beta}_t|\boldsymbol{y}_{1:t}) d\boldsymbol{\beta}_t\right) \quad . \tag{4.51}$$

Applying a normalisation condition, the expressions become equalities, whereas the initial specification is described as the probability density expectation given by $\prod_j q(w_{j,t}|\boldsymbol{y}_{1:t-1})$ which is in turn translated to the following definition:

$$q(\boldsymbol{\beta}_t|\boldsymbol{y}_{1:t}) \propto exp\left(\mathbb{E}_{q(\boldsymbol{w}_t|\boldsymbol{y}_{1:t})}\left(\log p(y_t, \boldsymbol{\beta}_t, \boldsymbol{w}_t|\boldsymbol{y}_{1:t-1})\right)\right) \tag{4.52}$$

$$= exp\left(\mathbb{E}_{q(\boldsymbol{w}_t|\boldsymbol{y}_{1:t})}\left(\log[p(y_t|\boldsymbol{\beta}_t, \boldsymbol{w}_t)\right)\left(p(\boldsymbol{\beta}_t|\boldsymbol{y}_{1:t-1})p(\boldsymbol{w}_t|\boldsymbol{y}_{1:t-1})\right]\right)\right) \tag{4.53}$$

$$= p(y_t|\boldsymbol{\beta}_t, \boldsymbol{w}_t) exp\left(\mathbb{E}_{q(\boldsymbol{w}_t|\boldsymbol{y}_{1:t})}(\log p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}) + \log q(\boldsymbol{\beta}_t + \boldsymbol{y}_{1:t-1}))\right) , \tag{4.54}$$

where $\boldsymbol{\beta}_t$'s t time prior realised by the previous period posterior $q(\boldsymbol{\beta}_{t-1}|\boldsymbol{y}_{1:t-1})$ via the recursive Kalman filter is expressed by $q(\boldsymbol{\beta_t}|\boldsymbol{y_{1:t-1}})$.

4.4.4 Inclusion of Stochastic Volatility

A growing body of recent literature emphasises the significance of time-varying volatility incorporation for enhancing both point and density forecasts Clark & Ravazzolo (2015) as neither the assumption of a known regression variance nor the implementation of an unknown, invariable time variance are accurate representations of macroeconomic time-series data. Thus, the examined model specification accommodates the evaluation of σ_t^2 in the TVP framework. Great emphasis should also been placed on the fact that regardless of the explicit time-series specification regarding the evolution of the stochastic volatility parameter, according to empirical evidence, no statistically significant discrepancies arise via the implementation of alternative macroeconomic volatility specifications in the case where the main examination interest lies in forecasting y_t. Hence, the variance discounting concept originally developed by West & Harrison (2006) and further developed by Rockova et al. (2020) is embodied in the model specification. Defining the inverse variance as $\phi_t = \frac{1}{\sigma_t^2}$, employing the assumption of West & Harrison (2006), the conjugate form of the $t-1$ ϕ posterior is denoted by:

$$\phi_{t-1}|y_{1:t-1} \sim Gamma(\alpha_{t-1}, \beta_{t-1}). \tag{4.55}$$

Furthermore, for the conjugacy maintenance, a time prior is initialised as follows:

$$\phi_t|y_{1:t-1} \sim Gamma(\delta\alpha_{t-1}, \delta b_{t-1}) \ , \tag{4.56}$$

where δ denotes the variance discounting factor ranging between $0 < \delta < 1$ relevant to a hyperparameter set α_0, b_0. Hence, the t posterior mean of ϕ_t based on the outlined regime is given by:

$$\hat{\phi}_t = \mathbb{E}_{q(\beta_t|y_{1:T})}(\phi_t|y_{1:t}) = \frac{\alpha_t}{b_t} \ , \tag{4.57}$$

with $\alpha_t = 1/2 + \delta\alpha_{t-1}$ and $d_t = \frac{1}{2}\left[(y_t - \boldsymbol{x_t m_{t|T}})^2 + \boldsymbol{x_t P_{t|T} x_t'}\right] + \delta b_{t-1}$ where $\boldsymbol{m_{t|T}}$ and $\boldsymbol{P_{t|T}}$ signify the smoothed first two moments of β_t respectively. Put differently, the outlined framework introduces the exponential discounting of the past data information determined by the prior hyperparameter δ which attributes the relative weight to old versus recent realisations. In the special case where $\delta = 1$ a conventional recursive update regime is implemented for the approximation of the posterior whilst for a quicker time evolution of system's precision the assigned value ranges between $0.99 - 0.8$, according to empirical suggestions. Accounting for the evolution features of the investigated macroeconomic time-series data characterised by precipitous breaks in and leptokurtosis in periods of recession, the variances are expected undergo rapid movements with time, thus the initialised hyperparameter value is $\delta = 0.8$.

Observing the form of the outlined formulas one can obtain the counter-cyclical iterative update regime of ϕ_t subject to the evaluation of ϕ_{t-1}. A backward recursive filter is applied in the examined algorithm smoothing the ϕ_t evaluations.

Specifically, the backward recursive filter originally introduced by West & Harrison (2006) is expressed by specification 31 in Algorithm 2, with $t = T-1$, $\hat{\phi}_T = \tilde{\phi}_T$ and $\tilde{\phi}_t = \mathbb{E}_{q(\beta_t|y_{1:t})}(\phi_t|y_{t-1})$ and once the precision of ϕ_t is evaluated the associated volatility posterior mean estimate is calculated by the inverse of $\tilde{\phi}_t$.

At last, the parametric update process followed by by the simulation study in the basis of the VB interference is accurately depicted in Algorithm 2, where the output signal obtained comprises of $\boldsymbol{m}_{t|T}$ realisations, signifying the smooth posterior of $q(\boldsymbol{\beta}_t|y_{1:T})$, from which the expectation of $\gamma_{i,j}$, $\pi_{0,t}$ and $\tau^2_{j,t}$ are approximated with respect to $q(\boldsymbol{\beta}_t|y_{1:T})$. Thus, based on the previous analysis the updating operations embedded in the dynamic variable prior regime subject to the VP posteriors are expressed by:

$$\hat{\tau}^2_{j,t} = \mathbb{E}[q(\tau^2_{j,t}|y_t)] = (h_0 + m_{j,t|T})/(g_0 + 1/2) \ , \tag{4.58}$$

$$\hat{\gamma}_{j,t} = \mathbb{E}[q(\gamma_{j,t}|y_t)] \tag{4.59}$$

$$= \frac{N\left(m_{j,t|T}|0, \hat{\tau}^2_{j,t}\right)\hat{\pi}_{0,t}}{N\left(m_{j,t|T}|0, \hat{\tau}^2_{j,t}\right)\hat{\pi}_{0,t} + N\left(m_{j,t|T|t}|0, \underline{c} \times \hat{\tau}^2_{j,t}\right)(1-\hat{\pi}_{0,t})} \ , \tag{4.60}$$

$$\hat{v}_{j,t} = \mathbb{E}[q(v_{j,t}|y_t)] = (1-\hat{\gamma}_{j,t})^2 \underline{c}\hat{\tau}_{j,t} + \hat{\gamma}_{j,t}\hat{\tau}^2_{j,t} \ , \tag{4.61}$$

$$\hat{\pi}_{0,t} = \mathbb{E}[q(\pi_{0,t}|y_t)] = \left(1 + \sum_{j=1}^{p} \hat{\gamma}_{j,t}\right) / (2+p) \ . \tag{4.62}$$

Additionally, based on Algorithm 2 one can observe that algorithmic specifications 4-12 denote the Kalman filter application in regards to the TVP state-space specification, 13-17 outline the backwards recursion process and similarly 28-33 apply the stochastic volatility updates.

Algorithm 2 Variational Bayes algorithm estimation of time varying parameter regression specification with stochastic variance and dynamic variable selection.

Initialization of all matrices and vectors $\boldsymbol{m_0}$,$\boldsymbol{P_0}$, α_0 , b_0 , $c_{j,0}$, $d_{j,0}$, g_0 , h_0 , $\underline{c}$ and δ .
$r = 1$
while $||\mathcal{G}(q(\boldsymbol{\beta}^{(r)}), \boldsymbol{w}^{(r)}|\boldsymbol{y}) - \mathcal{G}(q(\boldsymbol{\beta}^{(r-1)}), \boldsymbol{w}^{(r-1)}|\boldsymbol{y})|| \to 0$ **do**
  for $t = 1$ **to** T **do**
    $\tilde{\boldsymbol{W}}_t^{(r)} = diag\left(\left(w_{1,t}^{-1(r-1)} + v_{1,t}^{-1(r-1)}\right)^{-1}, .., \left(w_{p,t}^{-1(r-1)} + v_{p,t}^{-1(r-1)}\right)^{-1}\right)$
    $\tilde{\boldsymbol{F}}_t^{(r)} = \tilde{\boldsymbol{W}}_t^{(r)} \left(\boldsymbol{W}_t^{(r-1)}\right)^{-1}$
    $\boldsymbol{m}_{t|t-1}^{(r)} = \tilde{\boldsymbol{W}}_t^{(r)} \boldsymbol{m}_{t-1|t-1}^{(r)}$ // Predicted mean value
    $\boldsymbol{P}_{t|t-1}^{(r)} = \tilde{\boldsymbol{F}}_t^{(r)} \boldsymbol{P}_{t-1|t-1} \tilde{\boldsymbol{F}}_t^{(r)'} + \tilde{\boldsymbol{W}}_t^{(r)}$ // Predicted variance
    $\boldsymbol{K}_t^{(r)} = \boldsymbol{P}_{t|t-1}^{(r)} \boldsymbol{x}_t' \left(\boldsymbol{x}_t \boldsymbol{P}_{t|t-1}^{(r)} \boldsymbol{x}_t' + \hat{\sigma}^{2(r-1)}\right)^{-1}$ // Kalmal gain
    $\boldsymbol{m}_{t|t}^{(r)} = \boldsymbol{m}_{t|t-1}^{(r)} + \boldsymbol{K}_t^{(r)} \left(y_t - \boldsymbol{x}_t \boldsymbol{m}_{t|t-1}^{(r)}\right)$ // β_t filtered mean
    $\boldsymbol{P}_{t|t}^{(r)} = \left(\boldsymbol{I}_p - \boldsymbol{K}_t^{(r)} \boldsymbol{x}_t\right) \boldsymbol{P}_{t|t-1}^{(r)}$ // β_t filtered variance
  end for
  for $T = T - 1$ **to** 1 **do**
    $\boldsymbol{C} = \boldsymbol{P}_{t|t}^{(r)} \tilde{\boldsymbol{F}}_t^{(r)} \left(\boldsymbol{P}_{t+1|t}^{(r)}\right)^{-1}$
    $\boldsymbol{m}_{t|T}^{(r)} = \boldsymbol{m}_{t|t}^{(r)} + \boldsymbol{C}\left(\boldsymbol{m}_{t+1|T}^{(r)} - \boldsymbol{m}_{t+1|t}^{(r)}\right)$ // β_t smoothed mean value
    $\boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t} + \boldsymbol{C}\left(\boldsymbol{P}_{t+1|T}^{(r)} - \boldsymbol{P}_{t+1|t}^{(r)}\right)\boldsymbol{C}'$ // β_t smoothed variance
  end for
    $\boldsymbol{D}_t = \boldsymbol{P}_{t|T}^{(r)} + \boldsymbol{m}_{t|T}^{(r)} \boldsymbol{m}_{t|T}^{(r)'} + \left(\boldsymbol{P}_{t-1|T}^{(r)} + \boldsymbol{m}_{t-1|T}^{(r)} \boldsymbol{m}_{t-1|T}^{(r)'}\right)\left(I_p - 2\tilde{\mathcal{B}}_t^{(r)}\right)'$
    // State eqn squared error
    $R_t = \left[\left(y_t - \boldsymbol{x}_t \boldsymbol{m}_{t|T}^{(r)}\right)^2 + \boldsymbol{x}_t \boldsymbol{P}_{t|T} \boldsymbol{x}_t'\right]$ // Estim. eqn. squared error
    for $t = 1$ **to** T **do**
      for $j = 1$ **to** p **do**
      $\hat{\tau}_{j,t}^{-2(r)} = (g_0 + 0.5) / \left(h_0 + 0.5\left(m_{j,t|T}^{(r)}\right)^2\right)$ // Posterior mean of $\frac{1}{\tau_{j,t}^2}$
      $\hat{\gamma}_{j,t}^{(r)} = \frac{N\left(m_{j,t|T}^{(r)}|0, \hat{\tau}_{j,t}^{2(r)}\right)\hat{\pi}_{0,t}^{(r-1)}}{N\left(m_{j,t|T}^{(r)}|0, \hat{\tau}_{j,t}^{2(r)}\right)\hat{\pi}_{0,t}^{(r-1)} + N\left(m_{j,t|T|t}^{(r)}|0, \underline{c} \times \hat{\tau}_{j,t}^{2(r)}\right)\left(1 - \hat{\pi}_{0,t}^{(r-1)}\right)}$
      // Posterior mean of $\gamma_{j,t}$
      $\hat{v}_{j,t}^{r} = \left(1 - \hat{\gamma}_{j,t}^{(r)}\right)^2 \underline{c}\hat{\tau}_{j,t}^{2(r)} + \hat{\gamma}_{j,t}^{(r)} \hat{\tau}_{j,t}^{2(r)}$ // posterior mean of $v_{j,t}$
      $\hat{w}_{j,t}^{-1(r)} = (\underline{c}_0 + 0.5)/(\underline{d}_0 + 0.5\boldsymbol{D}_{jj,t})$ // Posterior mean of $\frac{1}{w_{j,t}}$
      end for
      $\hat{\pi}_{0,t}^{(r)} = \left(1 + \sum_{j=1}^{p} \hat{\gamma}_{j,t}^{(r)}\right)$ / (2+p) // Posterior mean of $\pi_{0,t}$
      $\hat{\phi}_t^{(r)} = (\delta\alpha_{t-1} + 0.5)/(\delta b_{t-1} + 0.5R_t)$ // Filtered mean of $\frac{1}{\sigma_t^2}$
    end for
    for $T = T - 1$ **to** 1 **do**
      $\tilde{\phi}_t^{(r)} = (1 - \delta)\hat{\phi}_t^{(r)} + \delta\tilde{\phi}_{t-1}^{(r)}$ // Smoothed mean of $\frac{1}{\sigma_t^2}$
    end for
  $r = r + 1$
end while

4.5 Additional Model Specifications Examined

A wide range of forecasting model specifications are employed by the current examination all of which are characterised as heavily parametrised TVP specifications with varying boundary conditions and prior hyperparameter selections. Most of the outlined models were firstly introduced by their respective authors for the inflation forecasting application and are selected based on their superior forecasting performance regarding the examined time-series. For reasons of clarity, all models' initialisation conditions strictly follow the "default" conditions as firstly introduced by the original literature paper.

4.5.1 KP-AR - Koop and Potter 2007 AR(p) Model With Structural Breaks

The examined model originally developed by Koop & Potter (2007), introduces a novel approach regarding change-point modelling. Its main advancement is that unlike previously excising models, the underlying regime does not necessitate the specification of the change point number in the data sample, but instead entails a Poisson probability mass function i.e. $d_i - 1 \sim Poisson(\lambda_i)$ for the regime duration distribution approximation d_i, nesting together both the general TVP regression specification with a change-break in every executive period along with a small regime number change-point model. In particular, the evolution of the coefficients is approximated as a random walk such that $\beta_j = \beta_{j-1} + u_j$, where $u_j \sim N_m(0, B_0)$ or according to the Bayesian regime the hierarchical prior is denoted by $\beta_j | \beta_{j-1} \sim N(\beta_{j-1}, B_0)$, establishing complete independence between consecutive regime coefficients. Thus the KP prior is expressed as follows:

$$\beta_j \sim N_m(\beta_{j-1}, B_0)\,, \quad \beta_0 \sim N_m(0, \underline{V}_\beta)\,, \quad B_0^{-1} \sim Wishmart(\underline{\xi}, \underline{B})\,, \tag{4.63}$$

$$\omega_j \sim N(\omega_{j-1}, \delta), \qquad \omega_0 \sim N(0, \underline{V}_\omega), \qquad \delta^{-1} \sim Gamma(\underline{\kappa_1}, \underline{\kappa_2}), \tag{4.64}$$

where m signifies the total element number in the dependent variable x_t, $\underline{\xi} = m+1$, $\omega_j = log\sigma_j$, where σ_j denotes the error variance parameter, $\underline{\mathrm{V}}_\beta = I_m$, I_m resembles the identity matrix, B_0 denotes the prior covariance and $\underline{\kappa_1} = \underline{\kappa_2} = 0.5$ denote the decay factors. Additionally, changes in both the conditional means and variances are also obtained via the implementation of a Markov Chain posterior sampler, such that numerical specification of the KP-AR model is defined as follows:

$$y_{t+h} = x_t \beta_{st} + \sigma_{st} \varepsilon_{t+h} \ , \tag{4.65}$$

$$\beta_{st} = \beta_{st-1} + \eta_{st} \ \ , \tag{4.66}$$

$$\log(\sigma_{st}) = \log \sigma_{st-1} + \zeta_{st} \ \ , \tag{4.67}$$

where $t = 1, 2, ..., T$, $\varepsilon_{t+h} \sim N(0, 1)$ denotes the error term, η_{st} and ζ_{st} denote the state errors as a mixture of normal components, x_t denotes a $K-$ dimensional vector including all relevant explanatory variables namely an intercept and in this case the two lagged dependent variables and $s_t \in \{1, ..., K\}$ indicates a K-sate Markov-switching operator. The current regime employs the distinguishing assumption of a block diagonal transition probability matrix enabling the permanent transition between regimes - structural breaks in contrast to mainstream regime-switching frameworks. Finally, following the structural-break literature contribution of Bauwens et al. (2015), where substantial attention has been attributed on the fabrication of optimal hierarchical prior estimation values in regards to each regime duration and characteristic parameter, the current study initialises a maximum state number $K_{max} = 10$ in combination with a Gibbs sampler for the appropriate quantification of structural-break number and similar initial conditions and geometric prior values are assumed in the restricted Markov regime for $S_T = (s_1, s_2, ..., s_T)'$:

$$Pr(s_t = i | s_{t-1} = i) = p_i \ \ , \tag{4.68}$$

$$Pr(s_t = i + 1 | s_{t-1} = i) = 1 - p_i \ \ . \tag{4.69}$$

4.5.2 UC-SV - Stock and Watson 2007

The current section introduces a prototype model originally developed by Stock & Watson (2007) known as Unobserved Component Stochastic Volatility (UC-SV) model. In this regime, inflation rate is approximated as the superposition of two unobserved components namely a serially uncorrelated transitory disturbance ε_t and a stochastic trend disturbance τ_t. Additionally, the evolution of the corresponding time-varying variance terms is approximated via logarithmic random walks locally specified as follows:

$$y_{t+h} = \tau_t + \sigma_t^{\varepsilon}\varepsilon_{t+h} \ , \qquad \varepsilon_t \sim N(0,1) \tag{4.70}$$

$$\tau_t = \tau_{t-1} + \sigma_t^{\eta}\eta_t \ , \qquad \eta \sim N(0,1) \tag{4.71}$$

$$log\sigma_t^{\varepsilon} = log\sigma_{t-1}^{\varepsilon} + \nu_t^{\varepsilon} \ , \qquad \nu_t^{\varepsilon} \sim N(0,\gamma_2) \tag{4.72}$$

$$log\sigma_t^{\eta} = log\sigma_{t-1}^{\eta} + \nu_t^{\eta} \ , \qquad \nu_t^{\eta} \sim N(0,\gamma_1) \tag{4.73}$$

where $t = 1, 2, ..., T$, γ_1 and γ_2 are scalar parameters that determine the smoothness of the evolutionary process regarding the evolution of stochastic volatility. According to literature the underlying model specification is characterised by both time-varying flexibility and parsimony not only in regards to inflation forecasting but also to other time-series due to its universality as the only required specifications include the volatility parameter scalar prior values and the initialised boundary conditions. Specifically, the hyperparameter selection aligns with the default conditions implemented in the Bauwens et al. (2015) study.

4.5.3 GK-AR - Giordani and Kohn 2008 AR(p) Model With Structural Breaks

Extending upon the work of Koop & Potter (2007), Giordani & Kohn (2008) introduced a mixture innovation structural-breaks model. Unlike multiple change point models' sampling schemes where the break dates are determined subject to all system parameters, most intrinsic parameters in this regime are classified as states and integrated out of the system enhancing sampling efficiency. Hence, the dynamic-mixture model is determined as follows:

$$y_{t+h} = x_t \beta_t + \varepsilon_{t+h} \ , \tag{4.74}$$

$$\beta_t = \beta_{t-1} + K_t \eta_t \ , \tag{4.75}$$

where $t = 1, 2, ..., T, \ K_t \in \{0, 1\}$ signifies an independent sequence Bernoulli indicator that determines whether a break is present at time t with probability π such that:

$$K_t = \begin{cases} 1 & \text{with probability } \pi \\ 0 & \text{with probability } 1 - \pi \ , \end{cases}$$

x_t is a $K-$ dimensional vector including all relevant explanatory variables, $\varepsilon_{t+h} \sim N(0, \sigma_t^2)$ denotes the innovation term and $\eta \sim N(0, Q)$ where Q signifies the state covariance matrix of size $p \times p$. According to literature additive outliers and structural breaks often give rise to high correlations between K and x resulting in highly inefficient sampling approaches. Therefore, the sampling algorithm adopted, originally developed by Gerlach et al. (2000), entails a conditional Gaussian approximation processes for drawing K unconditional of x. Specifically, a Gibbs sampling scheme is applied for the standard manner estimation of all system parameters' conditional posteriors which are then used for the assessment of K_t posterior sample. The posterior initialisation conditions employed follow the study of Bauwens et al. (2015).

4.5.4 Time Varying Dimension - TVD Model, Chan et al 2012

Expanding on the literature of dynamic mixture state-space models, Chan et al. (2012) developed a time varying dimension model that aims restriction deployment on the examined TVP regression for the elimination of errors associated with over-parametrisation challenges enabling at the same time the gradual evolution of system parameters expressed as follows:

$$y_{t+h} = \sum_{j=1}^{p} x_{j,t} s_{j,t} \beta_{j,t} + \varepsilon_{t+h} \ , \tag{4.76}$$

$$\beta_t = \beta_{t-1} + \eta_t \ \ , \tag{4.77}$$

where $t = 1, 2, ..., T, \ s_{j,t} \in 0, 1$ denotes a binary predictor indicator variable. To be precise, in the case where $s_{j,t} = 1$, then predictor j is embedded in the regression model at period t and similarly when $s_{j,t} = 0$, then the examined predictor is excluded form the system in period t. The underlying specification enables the readjustment of predictors included in the regression at each time period t, identifying the optimal predictor variable combination that most accurately quantify the impact on inflation dynamics ensuring parsimony. The boundary conditions and prior values adopted are in line with the suggested authors' exact implementation. The examined specification is equivalent to the first specification presented in the paper Chan et al. (2012) and for further details one can refer to the paper's online Appendix.

4.5.5 TVP-BMA Model, Groen, Paap and Ravazzolo 2013

The current model extends on the proposed model of Groen et al. (2013) where model uncertainty in terms of predictor variable combinations is incorporated, expressed by the following specification:

$$y_{t+h} = \sum_{j=1}^{p} x_{jt} s_j \beta_{jt} + \sigma_t \varepsilon_{t+h} \ , \tag{4.78}$$

$$\beta_t = \beta_{t-1} + \eta_t \ , \tag{4.79}$$

where $t = 1, 2, ..., T, \ \varepsilon_{t+h} \sim N(0, 1)$ indicates the measurement error attributing stochastic volatility and structural breaks are incorporated in both the regression variance and parameters by the varying nature of both β_{jt} and σ_t. Once again

$s_j \in \{0,1\}$ signifies a binary indicator introducing model uncertainty in the examined regime. Specifically, for a total number of p variables s_j denotes the inclusion of variable x_j in the regression specification for $j = 1, 2, ..., p$. Additionally, each s_j parameter's Bernoulli posterior is expressed as a sequence of binary numbers, hence the posterior mean signifies the precise inclusion probability in regards to each regression predictor j as often applied in the Bayesian Model Averaging processes. The current regime involves a Bernoulli prior with inclusion variable probabilities equal to 0.5 and all initialised boundary conditions and specification are in line with the study of Bauwens et al. (2015).

4.5.6 Time Varying Sparsity - TVS Model, Kalli and Griffin 2014

The dynamic regression model examined relates a response y_t to the regressor variables $x_{j,t}$ via the following specification originally introduced by Kalli & Griffin (2014):

$$y_{t+h} = \sum_{j=1}^{p} x_{j,t}\beta_{j,t} + \varepsilon_{t+h} \ , \tag{4.80}$$

$$\beta_{j,i} = (1-\alpha_j)_{j,t}\beta_{j,t-1} + \alpha_j \eta_{j,t} \ , \tag{4.81}$$

where $t = 1, 2, ..., T$, $\alpha_j \in [0,1]$ denotes a parameter that determines the temporal correlation between system variables, the hierarchical parameter $\psi_{j,t}$ indicates a first-order autoregressive gamma procedure such that $var(\eta_{j,t}) = \psi_{j,t}$ and $\rho_{j,t} = \sqrt{\frac{\psi_{j,t}}{\psi_{j,t-1}}}$ and determines the shrinkage of all square estimates towards zero quantifying the relevance of the i^{th} regressor such that a greater $\psi_{j,t}$ value signifies a greater relevance of the associated regressor at time t, as explained by both Pitt et al. (2002) and Pitt & Walker (2005). The current regime approximates the $\beta_{j,t}$ prior in a normal-gamma autoregressive manner that according to literature, normal-gamma priors in linear regression models display exceptional shrinkage properties. More specifically, in the outlined configuration each predictor's coefficient undergoes flexible shrinkage in some periods whilst in other remains unrestricted. The initialised boundary conditions are employed according to the original paper.

4.5.7 TVP-AR Model, Petenuzzo and Timmermann 2017

According to the findings of Pettenuzzo & Timmermann (2017), forecasting performance of the TVP-AR model seems to outperform a vast majority of similar models in regards to inflation in particular. The regression specification includes an intercept and two autoregressive lags with all regression variances and coefficients exhibiting time variation. The examined regime is a generalisation of the UC-SV model, allowing autoregressive terms in the model specification. Nevertheless, the forecasting equation only includes stochastic volatility features without the integration of any state equations. Finally, the adapted priors of the current examination are in line with the values suggested by Bauwens et al. (2015).

4.5.8 LASSO - The Bayesian Lasso Model

The approximation of linear regression parameters via the Lasso estimate can be thought as a similar Bayesian posterior regime estimate, where all independent variables are characterised by independent - double exponential Laplace priors. The structure of the hierarchical lasso prior is denoted as:

$$p(\beta|\sigma^2, \tau_1^2, \tau_2^2, .., \tau_p^2) \sim N(0, \sigma^2 V) \ , \tag{4.82}$$

$$p(\tau_j^2) \sim Exponential\left(\frac{\lambda^2}{2}\right) \ , \quad forj = 1, 2, ..., p \ , \tag{4.83}$$

$$p(\lambda^2) \sim Gamma(r, \delta) \ , \tag{4.84}$$

$$p(\sigma^2) \sim \frac{1}{\sigma^2} \ , \tag{4.85}$$

where $p(\sigma^2)$ represents the scale-invariant noninformative marginal prior of σ^2, λ^2 denotes a hyperprior that allocates higher density values at the marginal maximum likelihood estimate, $V = diag(\tau_1^2, \tau_2^2, .., \tau_p^2)$ and r, δ correspond to the shape and inverse scale parameter of the gamma distribution respectively. Following the authors suggestion the initialised conditions include $r = 1$ and $\delta = 3$. Finally, implementing a recursive sampling regime the posterior is evaluated subject to the conditional posteriors as follows:

$$\beta|\sigma^2, \{\tau_j^2\}_{j=1}^p, y \sim N\left((x'x + V^{-1})^{-1}x'y, \sigma^2(x'x + V^{-1})^{-1}\right) \ , \tag{4.86}$$

$$\frac{1}{\tau_j^2}|\beta, \sigma^2, y \sim IG\left(\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2\right) \ , for \ j = 1, 2, ..., p \ , \tag{4.87}$$

$$\lambda^2|\beta, \sigma^2, \{\tau_j^2\}_{j=1}^p, y \sim Gamma\left(p + r, \frac{1}{2}\sum_{j=1}^{p}\tau_j^2 + \delta\right) \ , \tag{4.88}$$

$$\sigma^2|\beta, \{\tau_j^2\}_{j=1}^p, y \sim iGamma\left(\frac{T-1}{2} + \frac{p}{2}, \frac{\Psi}{2} + \frac{1}{2}\beta'V^{-1}\beta\right) \ , \tag{4.89}$$

where IG is the acronym of Inverse Gaussian distribution and $\Psi \in R^{D\times D}$ denotes the covariance matrix.

4.5.9 Bayesian Model Averaging - BMA Model

Unlike all other models examined, the BMA specification entails a many predictor constant parameter regression expressed as follows:

$$y_{t+h} = x_t\beta + \varepsilon_{t+h} \tag{4.90}$$

where $t = 1, 2, ..., T$, σ^2 denotes the error variance term such that $\varepsilon \sim N(0, \sigma^2)$ and β signifies a $p \times 1$ vector encoding all time-varying parameters. The hierarchical prior implemented in the current specification regime includes the stochastic search variable selection (SSVS) approach expressed as follows:

$$p(\beta_i|\gamma_1) \sim (1-\gamma_i)N(0, \tau_0^2) + \gamma_i N(0, \tau_1^2) \ , \tag{4.91}$$

$$p(\gamma_i|\pi) \sim Bernoulli(\pi) \ , \tag{4.92}$$

where the stochastic variance σ^2 is represented by a diffusion prior and the parameter values include $\tau_0 = 0.001$, $\tau_1 = 4$ and $\pi = 0.5$. According to the examined priors, the posterior distribution subject to the recursive sampling of the conditional posterior is expressed by:

$$\beta|\sigma^2, \gamma, y \sim N\left((x'x/\sigma^2 + V^{-1})^{-1}x'y/\sigma^2, (x'x/\sigma^2) + V^{-1})^{-1}\right) \quad , \tag{4.93}$$

$$\gamma_i|\beta_i, \pi, y \sim Bernoulli\left(\frac{N(0|\beta_i, \tau_i^2)\pi}{N(0|\beta_i, \tau_0^2)(1-\pi) + N(0|\beta_i, \tau_1^2)\pi}\right) \quad , \tag{4.94}$$

$$\sigma^2|\beta, y \sim iGamma\left(\frac{T}{2}, \frac{1}{2}\sum_{t=1}^{T}(y_t - x_t\beta)^2\right) \quad , \tag{4.95}$$

where V denotes a diagonal matrix with assigned i^{th} element values of τ_0^2 and τ_1 for $\gamma_1 = 0$ and $\gamma_i = 1$ respectively. As a final remark, all aforementioned estimation models initially developed by their corresponding authors are simulated via a MCMC Gibbs sampler and are chosen based on their performance superiority regarding forecasting inflation.

5 | Many Predictors Inflation Forecasting

Following the immense literature regarding time-varying parameter macroeconomic models the fundamental objective of this paper includes inflation forecasting. According to ample evidence, inflation time-series experience very distinct and complex structural breaks, therefore TVP models are considered as a favourable tool for assessing both monthly and quarterly time-series with similar characteristics as per Stock & Watson (2007), Koop & Korobilis (2010), Chan et al. (2012), and Pettenuzzo & Timmermann (2017). Whilst the potential forecasting advances of significantly large data samples are supported by a variety of empirical evidence, the primary purpose of this investigation focuses on both the implementation and evaluation of a wide model range outlining the advantages of the adopted modeling methodologies.

The standard inflation forecasting model examined in this paper is consistent with the analysis of Stock & Watson (1999). In particular, the Phillips curve approximation employed is defined as:

$$\pi^h_{t+h} - \pi_t = \phi_0 + z_t\theta(L) + \Delta\pi_t\gamma(L) + e_{t+h} \ , \tag{5.1}$$

where the dependent variable on the left hand side signifies the change in inflation rate for less frequent time intervals relevant to the sampling frequency, $\theta(L)$ and $\gamma(L)$ denote the polynomials entailed in the lag operator L, P_t denotes the price level at period t expressed in an annualised rate, $\pi^h_t = log(\frac{P_t}{P_{t-h}})\frac{1200}{h}$ signifies the $h-$ period inflation as a function of P_t, also reported in an annualised rate such that $\pi_t \equiv \pi^1_t = 1200 \times \log\left(\frac{P_t}{P_{t-1}}\right)$. In their forecasting study, Stock & Watson (1999) specify inflation an $I(1)$ variable whilst stationarity is assumed for all exogenous variables $z_t \sim I(0)$. Additionally, adopting the model modifications imposed by Korobilis (2019), as per the original suggestion of Stock & Watson (1999), factor estimates f_t are implemented in the examined framework replacing the high-dimensional z_t parameter via the assessment of principal components. Moreover, enhancing components are introduced to the original forecasting equation for the improvement of the models' predictive performance - namely a stochastic volatility estimator in combination with additional TVPs. Hence, the finalised forecasting model is conventionally specified as:

$$\pi_{t+h}^h - \pi_t = \phi_{t,0} + f_t\theta_t(L) + \Delta\pi_t\gamma_t(L) + e_{t+h} \ , \tag{5.2}$$

where f_t indicated a factor vector with significantly lower dimensions and $e_t \sim N(0, \sigma_t^2)$. It is worth mentioning that although the evaluation process includes the inflation gap $\pi_{t+h}^h - \pi_t$, all estimated forecasts are relevant to $\pi_{t_h}^h$.

The factors employed for the estimation of Equation (5.2) include the first two lags of the dependent variable in combination with the initial 20 principal component approximations p regarding the f_t factors as well as the first two lags of the factor estimates. As previously mentioned, the methodology approach for the estimation a TVP regression model entails the integration of GAMP into Equation (5.2) using the following parametrisations:

$$y_t = \pi_{t+h}^h - \pi_t \ , \qquad x_t = [1, f_t, \Delta\pi_t] \ , \tag{5.3}$$

$$\beta_t = (\phi_{t,0}, \theta_t(L)', \gamma_t(L)')' \ , \qquad e_{t+h} = \varepsilon_t \ . \tag{5.4}$$

Adopting the outlined static regression and implementing all gathered observations, the empirical model consists of approximately 700 volatility estimates and 30,000 regression coefficients. Furthermore, the boundary condition specifications required to be initialised prior to the simulation study are the specification of two scalar prior hyperparameters. In the case of the SBL prior regime given in Equations (4.12) - (4.13) the determination of the hyperparameters values follow the condition $\underline{\alpha} = \underline{\mathrm{b}} = 1 \times 10^{-10}$. The underlying estimation approach resembles the TVP regression in Equation(4.10) and is referenced as TVP-GAMP.

Moving forward, the wide variety of competing regression models and algorithms presented in Section(4) are engaged by the following empirical forecasting application, achieving a meaningful comparison between high-dimensional predictor models with stochastic volatility components. Regardless of the different characteristics, hyperparamters and boundary conditions of each specification's coefficient evolution process, over time all models converge towards a special representation case of Equation(4.10). Additionally, the two inflation lags and intercept of the model specifications that implicate shrinkage priors namely - ***TVD, TVS, BMA, TVP-BMA*** and ***TVP-GAMP*** do not experience static or dynamic shrinkage as it solely applies to exogenous information from the macroeconomic indicators.

Forecast evaluations are performed for the time horizons $h = 1, 2, 3, 4, 6, 9, 12$ corresponding to one month, two months, one-quarter, 4 months, two-quarters, three-quarters and one year ahead respectively. The examined horizons are chosen for the evaluation of the models short-term and long-term predictive abilities as well as their relevance in monetary policy decisions. Additionally, the forecast evaluation approach follows the methodology of Bauwens et al. (2015), entailing both the evaluation of the predictive accuracy of the mean square forecast errors (MSFE) and the average predictive likelihood logarithmic estimator $ALPL$ both of which are averaged over the out-of-sample investigation period.

Particularly, for the evaluation of the $h-$ period ahead forecast associated with a specific model i and variable j, denoted by $h^0_{i,j,\tau+h}$, the out-of-sample forecast period τ is required to be initialised. Therefore, 50% of the data sample is defined as the initialised estimation period for the assessment of the out-of-sample rolling-window forecasts. Assuming $h = 1$ the corresponding estimation period translates to $1960M1 - 1988M6$ for the estimation of $1988M7$ forecast and an additional observation is compounded in the estimation sample every recursive period until the entire sample is exhausted, leading up to the total number of $\frac{696}{2} - h = 348 - h$ observations.

Furthermore, the corresponding predictive density of the $h-$ period ahead forecast denoted by $p(y_{i,j,\tau+h}|Data_\tau)$ is provided by the simulation algorithm for the time frame of $\tau = \tau_0, ..., T-1$, where $\tau_0 = 1988M6$. Denoting the estimated value of $y_{i,j,\tau+h}$ as $h^0_{i,j,\tau+h}$, the mean squared forecast error of model i and variable j is calculated by:

$$MSFE_{i,h} = \frac{\sum_{\tau=\tau_0}^{\tau-h}[y^0_{\tau+h} - E(y_{i,j,t+h}|Data_\tau)]^2}{T - \tau_0 - h + 1} \tag{5.5}$$

$$= \frac{\sum_{\tau=\tau_0}^{\tau-h} e^2_{i,j,\tau+h}}{T - \tau_0 - h + 1} \ , \tag{5.6}$$

and for the meaningful quantification of the obtained results, all performance estimates regarding each individual model i are presented in comparison to the benchmark $AR(2)$ with an intercept, in a ratio form denoted as:

$$MSFE_{ijh} = \frac{\sum_{\tau=\tau_0}^{\tau-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\tau_0}^{\tau-h} e^2_{bcmk,j,\tau+h}} \tag{5.7}$$

where $e^2_{bcmk,j,\tau+h}$ and $e^2_{i,j,\tau+h}$ indicate the squared forecast errors in regards to the variable j for time τ and the forecast horizon h for the benckmark model and in relation to model i, $i \in$ *{KP-AR, GK-AR, TVP-AR, UCSV, TVD, TVS, TVP-BMA, BMA, TVP-GAMP}* respectively. It is standard practice in macroeconomic time-series forecasting to employ an AR(2) model specification for the benchmark representation as it outperforms a random walk specification. Lastly, observing the outlined expression, estimation values less than unity denote a superior forecasting performance relative to the benchmark model specification and vice-versa.

In addition, for a more comprehensive evaluation the utilisation of the forecasts predictive likelihoods is further motivated. Unlike $MSFE$ estimations which solely depend on the point forecast values disregarding the remaining predictive distribution, the entire predictive density of the forecast performance is assessed instead. Specifically, the predictive likelihood is determined by the evaluation of the predictive density subject to the obtained forecast outcome $y^0_{\tau+h}$. Thus, following the approach of Geweke & Amisano (2010), the differential of the average logarithmic predictive likelihood (ALPL), between the examined model i and the benchmark $AR(2)$ model is estimated by:

$$ALPL_{i,j,h} = \frac{1}{T - \tau_0 - h + 1} \sum_{\tau=\tau_0}^{\tau-h} (LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h}) \ , \tag{5.8}$$

where $bcmk$ is an acronym for benchmark model and $LPL_{i,j,\tau+h}$ signifies the predictive score in logarithmic form of variable j regarding model i estimated at $\tau+h$. Negative $ALPL_{i,j,h}$ values suggest that for forecast horizon h and variable j, the density forecasts of model i are less accurate compared with the density forecasts generated for the benchmark model specification and vice-versa.

In addition, the statistical significance of both point and density forecast differences is evaluated via an equal predictive accuracy pairwise test originally introduced by Diebold & Mariano (1995) assessing the reliability of the performance estimates - $MSFE$ and $ALPL$. This is achieved via the employment of standard critical values for the test conduction which accordingly depend on the two sided test with null hypothesis being the $AR(2)$ benchmark model. The statistical significance of each result is presented in the Result section tables with significance degrees $1\%, 5\%$ and 10% denoted by $***$, $**$ and $*$ respectively.

Finally, the initial simulation approach regarding all examination models entails a Gibbs MCMC sampler and the interference process entails a total number of 1500 posterior samples subject to a burn-in period of 500 samples giving rise to a total number of 2000 iterations. Due to the nature of the investigated forecasting application, the employed iteration number regardless of its small value fulfills the convergence criteria and accordingly once convergence is met, estimation values of the two posterior moments are obtained for further interference.

The quarterly averages of the data are obtained for further inflation forecasting using different frequency data processed via the dynamic variable selection specification. Hence, the resulting data set consists of 174 observations per explanatory variable ranging between $1960Q1 - 2017Q4$ and the forecasted series of interest include CPI and PCEPI. However, the transformation formula of the examined frequency in regards to the $h-$ quarter ahead forecast realisation is denoted by:

$$h_{t+h} = \left(\frac{400}{h}\right) ln\left(P_{t+h} - P_t\right) \quad , \tag{5.9}$$

and the transformations of the predictors comply with the standard literature norms. Additionally, the forecasting specification describing the $h-$ period ahead inflation estimate is denoted by the following expression:

$$y_{t+h} = \alpha_t + \phi_{1,t} y_t + \phi_{2,t} y_{t-1} + \boldsymbol{x}_t \boldsymbol{\beta}_t + \varepsilon_{t+h} \quad , \tag{5.10}$$

where α_t signifies the AR(2) regression intercept, y_t and y_{t-1} denote the autoregressive terms and $\boldsymbol{x_t \beta_t}$ the component and coefficient of the exogenous predictors included in the specification. Aiming for a more comprehensive analysis, a variety of alternative forecasting models is further engaged by the empirical application covering a broad spectrum of forecasting specifications. All implemented models are selected due to both their simplicity and replicability and comply with the inflation specification. Furthermore, due to their different properties, some of the boundary conditions assume invariant coefficients which in these cases the following normalisation conditions are implemented $\alpha_t = \alpha, \phi_{j,t} = 1 \forall j \in 1, 2..p$.

Moreover, the meaningfulness of the predictors sample size is evaluated by controlling for varying predictor subsets and combinations namely - (i) models with no predictor terms, (ii)10 predictor model including the first 10 explanatory variables, (iii) 50 predictors model including the first 50 explanatory variables (iv) all 129 principal components signifying all explanatory variables predictors. Each model category is illustrated by the following overview of model specifications and forecast metrics:

- **AR(2)**: OLS specification estimating the benchmark AR(2) model with an intercept.
- **TVPAR**: AR model specification with time varying parameters and a stochastic volatility component evaluated via MCMC iterations.
- **10 Factor model, FAC10**: A further extension of the benchmark model achieved via the augmentation of the 10 initial principal components extracted using OLS for the parsimonious portrayal of the relevant data set information.
- **10 Factor BAG model, BAG/FAC10**: Constant parameter regression with 10 predictors evaluated via the implementation of the bagging algorithm of Breiman (1999) controlling for the optimal selection of the most relevant factors in a static regime.
- **10 Factor DMA model, DMA/FAC10**: TVP regression specification implementing the aforementioned dynamic variable selection prior.
- **10 Factor VBDVS model, VBDVS/FAC10**: TVP regression specification with 10 predictors evaluated via a Variational Bayes variable selection prior.
- **10 Factor GPR model, GPR/FAC10**: A Gaussian Process Regression specification (GPR) with 10 predictors. Generally GPR demonstrates a resilient non-parametric approach used for the determination of the optimal inflation specification among the time-varying parameter specification of a more complex nonlinear specification.
- **50 Factor SSVS model, SSVS/FAC50**: An application of the SSVS prior via the implementation of MCMC for the evaluation of the augmented AR

specification using the initial 50 principal components which can be seen as a static representation of a VBDVS algorithm.

- **50 Factor ELN model, ELN/FAC50**: A constant parameter regression estimation encoding the initial 50 principal components evaluated via the implementation of the Elastic Net Algorithm (ELN) developed by Zou & Hastie (2005). The underlying specification resembles a widely used penalised likelihood estimator applied in high-dimensional data-sets.

- **50 Factor VBDVS model, VBDVS/FAC50**: TVP regression specification with 50 predictors evaluated via a Variational Bayes variable selection prior.

- **129 Factor ELN model**: A constant variable regression specification with all 129 variable predictors approximated via the employment of ELN regime.

- **129 Factor PLS model, PLS/FAC129**: Constant parameter least squared (PLS) regression encoding all 129 sample predictors. The econometric methodology is aligned with the principal component analysis. However, the factor extraction process differs as it solely depends on the independent variable.

- **129 Factor VBDVS model, VBDVS/FAC129**: TVP regression specification encoding all 129 predictors estimated via the implementation of the outlined Variational Bayes Dynamic Variable Selection prior.

Standard default boundary conditions are applied for all models based on the original suggestions of their respective authors. Rolling window forecast evaluations are performed for the time horizons $h = 1, 2, 4, 8$ quarters ahead for which 50% of the sample data are used for the formation of the in-sample evaluation period. In addition, the forecast evaluation analysis is in line with the empirical application of the monthly dataset such that the forecast accuracy is quantified via $MSFE$ measurements and accordingly the corresponding forecast densities are evaluates using $ALPL$ estimations.

6 | Forecasting Results

The average performance rankings of each model for every evaluation period are presented by Tables 6.1 to 6.10. Furthermore, for the identification of the underlying principles contributing to each model's performance both the cumulative sum of mean squared forecast errors (CSMSFE) and cumulative average logarithmic predictive likelihood differentials (CSAPLD) are graphically illustrated by Figures 4. 1- 4.6 providing evidence in regards to the forecasting gains, where:

$$CSMSFE_{iht} = \sum_{\tau=\tau_0}^{\tau-h} MSFE_{i,\tau+h} \tag{6.1}$$

$$CSALPLD_{iht} = \sum_{\tau=\tau_0}^{\tau-h} (ALPL_{i,\tau+h} - ALPL_{bcmk,\tau+h}) \tag{6.2}$$

The x-axis of the graphical illustrations represent the forecasting evaluation time period dates and similarly the y-axis indicate the associated CMSFE and CSAPLD estimations in regards to each examined model.

6.1 Monthly Forecasting Results

Table 6.1: Performance evaluation of CPI point forecasts: relativistic MSFE in regards to the benchmark AR(2) model using specification Equation (5.1).

	CPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*0.986*	*0.995*	*0.998*	*1.002*	*0.993*	*1.014*	*1.012*
UC-SV	*1.045**	*1.013***	*1.112***	*1.006****	*1.069****	*1.150****	*1.046**
GK-AR	*0.979*	*0.997*	*1.008*	*0.987*	*1.008*	*1.014*	*0.997*
TVD	*1.187***	*1.572*	*2.343*	*2.407**	*2.168*	*2.061**	*1.705*
TVS	*1.021*	*1.221*	*0.998*	*0.964*	*0.946*	*0.932*	*0.915*
BMA	*1.176***	*0.075****	*1.065****	*0.964*	*1.107*	*0.914****	*0.906****
TVP-BMA	*0.972*	*1.017*	*1.003*	*1.008*	*0.971*	*0.855*	*0.824*
TVP-AR	*1.103***	*1.189***	*1.097**	*1.042****	*1.006****	*0.902***	*1.017****
TVP-GAMP	*1.046***	*1.029***	*1.005****	*0.995*	*0.914**	*0.853***	*0.801****

Table 6.2: Performance evaluation of CPI point forecasts: relativistic MSFE in regards to the benchmark AR(2) model using specification Equation (5.2).

	CPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*1.061*	*1.003*	*0.714*	*0.893*	*0.785*	*0.672**	*0.538***
UC-SV	*0.978***	*1.013**	*1.174****	*1.221***	*1.243***	*1.178*	*1.321****
GK-AR	*0.989*	*1.003*	*0.996*	*0.998*	*0.874*	*0.886*	*0.892**
TVD	*1.014*	*0.927*	*0.889*	*0.989*	*0.857*	*0.863*	*0.841*
TVS	*0.931***	*0.935***	*0.963*	*0.976****	*0.897***	*0.854**	*0.924***
BMA	*0.985**	*0.993**	*0.948*	*0.890*	*0.892*	*0.714**	*0.760*
TVP-BMA	*0.922*	*0.916*	*0.918*	*0.897*	*0.834*	*0.804*	*0.663*
TVP-AR	*0.856***	*0.918***	*0.925****	*0.958***	*0.803****	*0.702****	*0.645****
TVP-GAMP	*0.933***	*0.906***	*0.817**	*0.832*	*0.813****	*0.754***	*0.772****

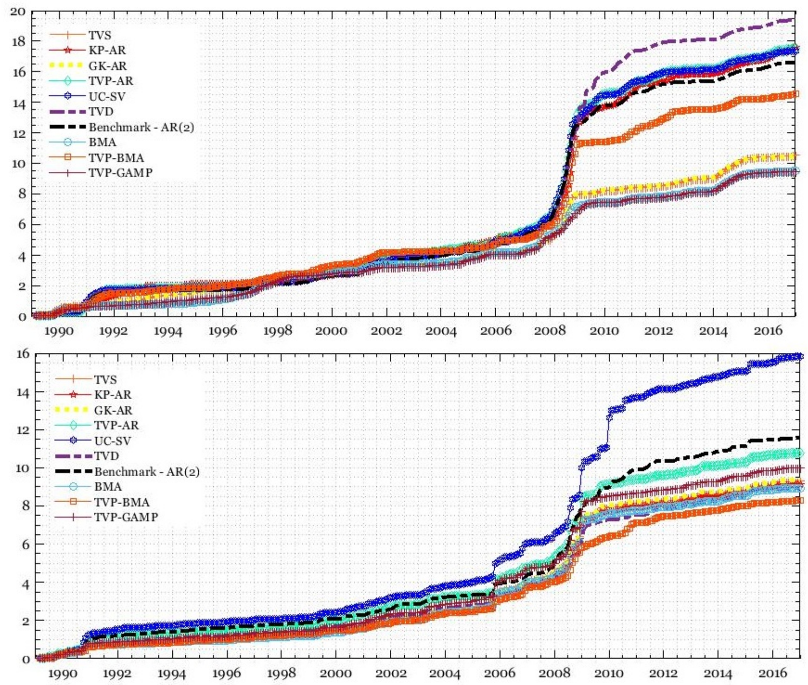


Figure 6.1: Graphical representation of CPI cumulative sum of monthly squared forecast errors for the evaluation period of $h = 12$, where the top illustration is based on specification Equation (5.1) and the bottom illustration is according to specification Equation (5.2).

Table 6.3: Performance evaluation of PCEPI point forecasts: relativistic MSFE in regards to the benchmark AR(2) model using specification Equation (5.1).

	PCEPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*0.998*	*0.975**	*0.935*	*0.971*	*1.100*	*0.901*	*0.994*
UC-SV	*1.044***	*1.053**	*1.014****	*0.955***	*0.935***	*1.054****	*1.008****
GK-AR	*1.007*	*0.997*	*0.990*	*1.004*	*1.007*	*1.042*	*0.989*
TVD	*0.926*	*1.137*	*1.501*	*2.596*	*1.660*	*2.109*	*1.662*
TVS	*1.039**	*1.338***	*1.412**	*1.291***	*1.177***	*1.061****	*0.991***
BMA	*0.980*	*1.006**	*1.015***	*1.088*	*1.038*	*0.949*	*0.918*
TVP-BMA	*1.037*	*1.061*	*1.220*	*1.002*	*0.969*	*0.973*	*0.908*
TVP-AR	*1.074****	*1.065****	*1.114**	*0.947***	*0.938****	*0.952****	*0.989****
TVP-GAMP	*1.064*	*1.058****	*1.014****	*1.026****	*1.082***	*1.028****	*0.967****

Table 6.4: Performance evaluation of PCEPI point forecasts: relativistic MSFE in regards to the benchmark AR(2) model using specification Equation (5.2).

	PCPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*1.024*	*1.007**	*1.053**	*1.027*	*1.149*	*0.989*	*0.972*
UC-SV	*1.033***	*1.004***	*1.082****	*1.071***	*0.981****	*0.962****	*1.003**
GK-AR	*1.002*	*0.988*	*0.986*	*0.996*	*0.992*	*1.101*	*0.983*
TVD	*1.197*	*1.169*	*1.205*	*1.227*	*1.299*	*1.321*	*1.232*
TVS	*1.219**	*1.158***	*1.042***	*1.002****	*0.994***	*0.971***	*0.969****
BMA	*1.022*	*0.952**	*0.915*	*0.876*	*0.851*	*0.702**	*0.711*
TVP-BMA	*1.016*	*0.961*	*0.936*	*0.956*	*0.906*	*0.718*	*0.706*
TVP-AR	*1.015***	*0.987****	*0.845**	*0.905***	*0.874****	*0.813****	*0.762***
TVP-GAMP	*0.935****	*0.958***	*0.936****	*0.837****	*0.856****	*0.764****	*0.753***

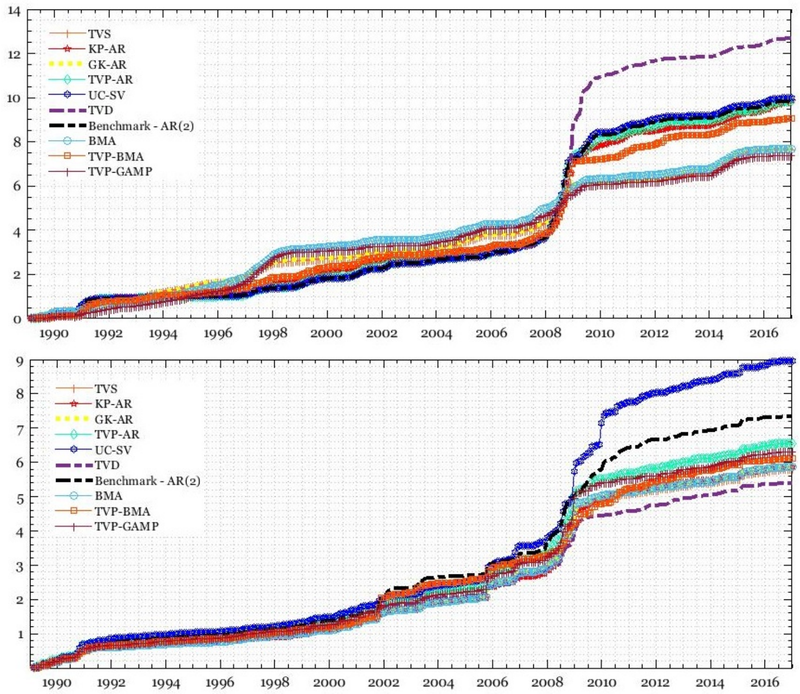


Figure 6.2: Graphical representation of PCEPI cumulative sum of monthly forecast errors for the evaluation period of $h = 12$, where the top illustration is based on specification Equation (5.1) and the bottom illustration is according to specification Equation (5.2).

Table 6.5: Performance evaluation of CPI density forecasts: relativistic ALPL in regards to the benchmark AR(2) model using specification Equation (5.1).

	CPI						
	h=1	*h=2*	*h=3*	*h=4*	*h=6*	*h=9*	*h=12*
KP-AR	*0.052*	*0.005*	*0.034*	*0.055*	*-0.092*	*0.014*	*0.126*
UC-SV	*-1.215*	*-1.339*	*-1.378**	*-1.484*	*-1.152**	*-1.177*	*-1.087**
GK-AR	*0.086***	*0.007*	*0.006*	*0.227*	*0.207*	*-0.140*	*-0.181*
TVD	*-0.056*	*-1.789*	*-0.259*	*-0.398*	*-0.066*	*-0.457*	*-0.342*
TVS	*0.013*	*-0.133*	*-0.089*	*-0.004*	*-0.028*	*-0.052*	*-0.019*
BMA	*0.065*	*0.068*	*-0.011*	*0.121***	*-0.048*	*0.212*	*-0.066*
TVP-BMA	*0.179**	*0.159*	*0.014*	*0.271*	*0.153*	*0.183*	*-0.033*
TVP-AR	*-0.06*	*-0.271*	*-0.308*	*-0.561*	*-0.762*	*0.271*	*-0.395*
TVP-GAMP	*0.027*	*0.092*	*0.103*	*0.362*	*0.427*	*0.321***	*0.187*

Table 6.6: Performance evaluation of CPI density forecasts: relativistic ALPL in regards to the benchmark AR(2) model using specification Equation (5.2).

	CPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*0.085*	*0.111*	*0.094*	*0.050*	*0.170*	*-0.674*	*0.103*
UC-SV	*-0.023***	*-0.026**	*-0.111*	*0.028*	*-0.094*	*0.007*	*-0.063***
GK-AR	*0.047****	*0.066*	*0.129***	*0.078**	*0.042*	*0.014**	*0.004*
TVD	*-0.061*	*-0.028*	*-0.034*	*-0.091*	*-0.010*	*-0.075*	*-0.024*
TVS	*-0.003*	*-0.017*	*0.014*	*-0.198*	*-0.053*	*-0.052*	*-0.158*
BMA	*0.066*	*0.059***	*0.025*	*0.071***	*-0.167*	*0.212*	*0.008*
TVP-BMA	*0.013*	*0.184*	*0.284**	*0.226*	*0.336*	*-0.183*	*1.310*
TVP-AR	*0.017*	*0.132*	*0.186*	*0.085*	*0.152*	*-0.280*	*0.100*
TVP-GAMP	*0.045*	*0.043****	*-0.014*	*0.121*	*0.024**	*0.032****	*0.027*

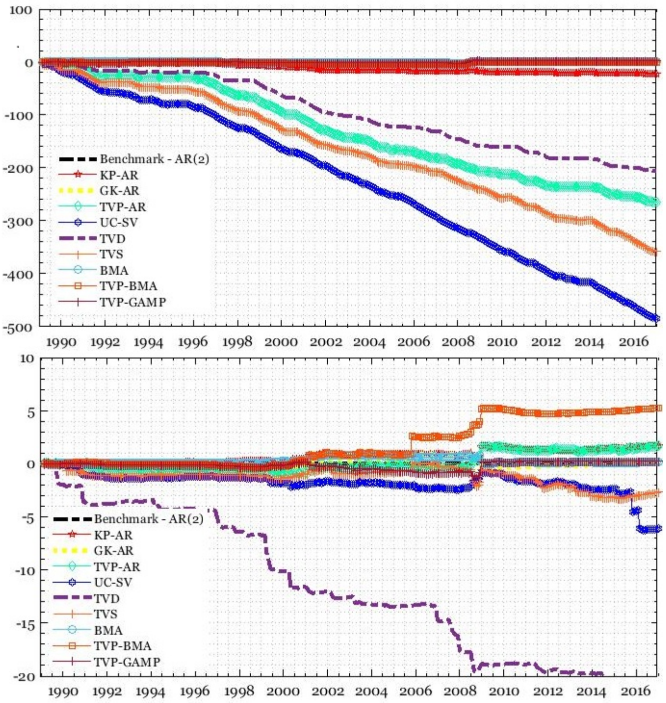


Figure 6.3: Graphical representation of CPI cumulative sum of logarithmic predictive likelihoods estimates difference between model i and the benchmark $AR(2)$ model for the evaluation period of $h = 12$, where the top illustration is based on specification Equation (5.1) and the bottom illustration is according to specification Equation (5.2).

Table 6.7: Performance evaluation of PCEPI density forecasts: relativistic ALPL in regards to the benchmark AR(2) model using specification Equation (5.1).

	PCEPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*0.026*	*0.018*	*0.004*	*0.005*	*0.043*	*0.043*	*0.104*
UC-SV	*-0.037*	*-1.238*	*-1.315*	*-1.432*	*-1.257*	*-1.103*	*-1.015*
GK-AR	*0.042*	*0.055*	*0.068*	*0.127*	*0.051**	*0.047**	*0.079*
TVD	*-0.292*	*-0.339*	*-0.598*	*-0.635*	*-0.584*	*-0.498*	*-0.531*
TVS	*-0.037*	*-0.328*	*-0.812*	*-0.916*	*-0.922*	*-0.964*	*-0.987*
BMA	*0.002*	*0.009*	*-0.003*	*0.069*	*-0.002*	*0.016*	*0.066*
TVP-BMA	*0.009*	*0.010*	*0.026*	*0.052*	*0.091*	*0.007*	*0.009*
TVP-AR	*-0.051*	*-0.702*	*-0.653***	*-0.713*	*-0.781*	*-0.819***	*-0.648****
TVP-GAMP	*0.056*	*0.015**	*0.073****	*0.016***	*0.025****	*0.014*	*0.048***

Table 6.8: Performance evaluation of PCEPI density forecasts: relativistic ALPL in regards to the benchmark AR(2) model using specification Equation (5.2).

	PCEPI						
	h=1	***h=2***	***h=3***	***h=4***	***h=6***	***h=9***	***h=12***
KP-AR	*0.045*	*0.066*	*0.005*	*-0.005*	*0.003*	*0.013*	*0.018*
UC-SV	*0.022**	*0.052*	*-1.076*	*-1.124*	*-1.043*	*-1.158*	*-1.689*
GK-AR	*-0.021*	*0.022***	*-0.064*	*0.236*	*-0.129*	*0.018*	*0.146**
TVD	*-0.826*	*-0.834*	*-1.734*	*-2.509*	*-4.574*	*-7.283*	*-7.259*
TVS	*0.067*	*-0.043*	*-0.511*	*-0.989*	*-1.165*	*-0.378*	*-0.892*
BMA	*0.044*	*0.029*	*-0.004*	*0.040*	*-1.159*	*-0.176*	*0.059*
TVP-BMA	*0.060*	*0.092*	*0.169*	*0.226*	*0.384*	*0.156*	*0.071*
TVP-AR	*0.045**	*0.347*	*0.382*	*0.520*	*0.061*	*0.158*	*0.289*
TVP-GAMP	*-0.087*	*0.033**	*0.0173*	*0.253****	*0.328*	*0.468***	*0.309**

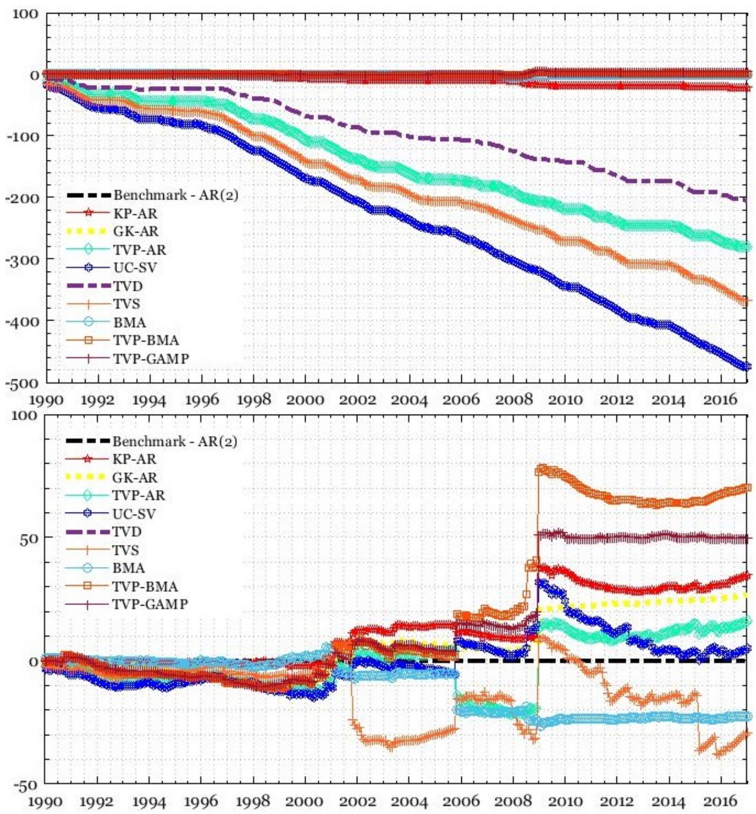


Figure 6.4: Graphical representation of PCEPI cumulative sum of logarithmic predictive likelihoods estimates difference between model i and the benchmark $AR(2)$ model for the evaluation period of $h = 12$, where the top illustration is based on specification Equation (5.1) and the bottom illustration is according to specification Equation (5.2).

6.1.1 Examining Inflation Forecast Gains For Monthly Data

Examining the monthly average point forecast estimates, Table 6.1 illustrates that under the initial regression specification regime denoted by Equaiton(5.1), the models outperforming the benchmark model for $h = 12$ are namely - GK-AR, TVS, BMA, TVP-BMA, and TVP GAMP, as further illustrated by the top panel of both Figure 6.1 and Figure 6.2. In particular, the TVP-GAMP point forecasts appear to dominate all other competing model specifications both in regards to CPI and PCEPI forecast evaluations, denoting increasing gains for longer time horizons.

Generally, the forecasting performances of models entailing larger predictor number namely - BMA, TVP-BMA and TVP-GAMP, appear to significantly improve with increasing time horizons compared to time-varying parameter models containing no predictors. Hence, this observation suggests that regardless of the evolution process approximation of the time-varying coefficients, the information enclosed in the explanatory predictors disposes greater relevance to the investigation objective of the current study. The underlying statement is not compromised by the statistical insignificance of TVP-BMA point forecasts or the inferior density forecasts of both BMA and TVP-BMA in respect to the TVP-GAMP density forecast performance.

Moreover, in the present regression specification, the absence of a stochastic volatility term in the BMA regime gives rise to model specification discrepancies, resulting in inferior density forecasts compared to the TVP-BMA and TVP-GAMP models which allow for variance fluctuations. In addition, since shrinkage in the TVP-BMA specification solely applies to predictors and no restrictions are imposed to the variation magnitude of time-varying parameters, the examined models specification is classified as overparameterized. Consequently, the associated point forecast performance is substandard relevant to the more inclusive GAMP model. Nevertheless, comparing BMA based specifications with the competing forecasting models, one identifies that overall, model averaging approximations demonstrate a dominant forecasting performance as opposed to model selection processes with time varying parameters. Particularly, performing a closer inspection at the behaviour of the model selection specifications, evidence outlines the highly volatile model size variations, with explanatory variables being included and excluded continuously between adjacent evaluation periods giving rise to continuous variations in the models composition which accordingly lead to severe interference implications and pervasive instability in both the parameter components and error variances when time-variation is embedded in model specifications. Lastly, the poorest forecasting performance obtained in the examined regime is associated with the TVD model, which is classified as inferior to TVP models with no predictors both in the cases of CPI and PCEPI time-series. In particular, the TVD estimation process is not only computationally challenging but there is also a distinct deterioration of its forecasting performance over longer evaluation horizons, denoting over-fitting challenges.

Hence, according to the empirical findings, the forecast performance improves via the implementation of a larger predictor number as opposed to the entailment of time-variation in both the parameters and volatilities. Increased forecasting gains are also obtained for model specifications encoding a stochastic volatility term and more inclusive model regimes imposing estimation constrains to the fluctuation magnitude of time-varying parameters. Furthermore, based on the empirical suggestions, the TVP-GAMP specification emerges the most suitable model specification due to its flexibility, ability to extract all relevant information enclosed in a large predictor set and incorporating an efficient structural break regime by regularising time variation and omitting time unnecessary predictor components. The realised results appear to contradict existing literature findings regarding TVP model specifications. A potential explanation of such occurrence is the delinquent assumption originally made by Stock & Watson (1999), describing inflation as an $I(1)$ variable, also denoted by specification Equation (5.1).

Once the dependent variable of the investigated regime is amended to the inflation gap, eliminating all random walk dynamics from the specification equation, the importance of time-varying parameters degenerates further and the predominant feature includes the information enclosed in the exogenous predictors. In sum, employing the alternative specification Equation (5.2), a tremendous advancement regarding all variant TVP model specifications is evident especially in the cases of TVD and TVP-AR approaches. Additionally, the aggregated range of both the CSMDFEs and CSALPLDs over the time evaluation period decrease considerably denoting a superior goodness of fit of the examined models in the alternative regime. The outlined statement is further supported by the bottom panel of Figure 6.1 and Figure 6.4 which highlights the superior performance of all examined models against the benchmark model, except UC-SV which yields the weakest forecasting performance in the examined model sample.

Upon closer examination of the point forecast realisations along with the corresponding MSFE results, differences emerge among various TVP specifications with greater emphasis being attributed to longer time horizons. The TVP-BMA regression specification demonstrates the minimum MSFE relative to the benchmark model among all specification models for $h = 12$ in regards to CPI forecasting and similarly the TVD specification for the PCEPI forecasts. Nonetheless, former empirical evidence underline the poor forecasting performance of the TVD model regarding the initial specification regime, latter findings rank the TVD regime amongst the best performing models due to its explicit feature of permitting for parameter variation and continuous modifications in the model dimensions over forecasting time. Therefore, the resulting point forecast evaluations of the TVD model exhibit extreme variation in alternative specification regimes which often give rise to performance spikes. The aforementioned phenomenon is reminiscent of the empirical findings of Groen et al. (2013). In addition both structural break models namely KP-AR and GK-AR collapse towards a more parsimonious representation of the alternative regime improving radically point forecast performance. However, the KP-AR model appears to outperform the GK-AR specification. A reasoning of this might be the Markov chain structure implemented for the approximation of the state sequence of KP-AR models generating less structural-breaks compared to GK-AR specifications which accordingly employ an independent Bernoulli prior. The feeblest estimation performance however, for both CPI and PECPI point forecast is associated with the UC-SV model specification. An explanation of the outlying finding includes the fact that the aforementioned specification approximates the evolution of variance logarithms as a Gaussian random walk, which gives rise to smoothness problems of the stochastic volatility in comparison to stochastic volatility specifications with heavier tails. Thus, in the cases where changes in data volatility are better approximated via a regime shift model, a smooth parameter variation model such as UC-SV, fails to capture any potential abrupt changes in volatility resulting from regime shifts leading to a poor forecasting performance. Although, in the initial specification regime, the TVP-GAMP model was classified as the best performing model, it appears to be outperformed by - KP-AR, BMA, TVP-BMA, TVD and TVP-AR specifications which exemplify in the alternative specification regime.

Inspecting the density forecast realisations, the results may not completely comply with the point forecast estimations but key features however are to some extent distinguishable. Adopting the initial specification regime, observing the top panel of Figure 6.3, the CSALPLDs portrays the substandard forecast accuracy of - UC-SV, TVS, TVP-AR, and TVD model specifications relative to the benchmark which appears to be further intensified with increasing time horizons, whilst the remaining models' density forecast performances are comparable to the benchmark model denoting reasonable predictive accuracy. Furthermore, conventional wisdom has it, that the incorporation of time variation is highly significant for the estimation of predictive densities as in highly volatile times no impact might be imposed on point forecasts but alternatively higher predictive estimates performance might deteriorate. Such wisdom is substantially reinforced by the realised findings. Table 6.5 and Table 6.7 indicate the superior density forecast performance of the specification models - TVP-GAMP and KP-AR compared with all remaining forecasting regimes in regards to CPI forecasts. In terms of the density forecasts of PCEPI, the models outperforming the benchmark model on average, include the KP-AR, GK-AR, BMA, TVP -BMA and TVP-GAMP denoting their enhanced resilience compared to CPI density forecasting. Due to the wide diversity of the examined model specifications in combination with the universality of each investigation regime, the task of completely identifying the most advanced econometric features contributing in point and density forecast evaluation for inflation forecasting is a rather trivial task.

Yet examining the alternative specification regime, the distinction between models forecasting accuracy becomes clearer and the substantial advancement of most model specifications is discernible. In particular, the models outperforming the benchmark specification outnumber the previous regression specification including - KP-AR, GK-AR, BMA, TVP-BMA, TVP -AR and TVP-GAMP. Additionally, the spread of CSALPLDs from the benchmark decreases tremendously, denoting the significance of parameter time-variation specification compared to the information enclosed in the exogenous predictors in regards to the forecasting regression (5.2). Overall, the empirical findings do not outline a single coherent narrative, but rather several narratives. Moreover, the empirical findings do not distinguish a single favourable forecasting method for the investigation objective as the forecasting performance of each model specification varies between alternative regression regimes.

As a final node, Figure 6.1 and Figure 6.2 identify crucial time events through the evaluation sample that trigger the diversification of models forecasting performance. Explicitly, the top panel of Figure 6.1 signifies the performance dominance of TVP-GAMP, closely followed by the BMA specification throughout the entire evaluation period which appear to intensify after the evaluation dates of middle

1990, 1999, 2005 and 2008 referring to inflation forecast of 2009. Similar models performance diversification events regarding evaluation dates are further denoted by the bottom panel of Figure 6.1 and Figure 6.2 reflecting the decisiveness of the aforementioned time periods. Hence, one concludes that regardless of the severe forecasting performance deterioration of all model specifications caused by the financial crisis, the underlying financial event does not represent the most crucial episode impacting the models forecasting performance but contrary additional "inflation-points" emerge in both early and late 1990s and 2005. Lastly, the greater divergence gap of forecasting performance is denoted approximately at the final 10 years of the out-of-sample evaluation period.

6.2 Quarterly Forecasting Results

Table 6.9, and Table 6.10 illustrate the MSFE and ALPL average results ranking for both CPI and PCEPI respectively. The results of the benchmark model are obtained by the statistical exercise whilst every other specification's estimate is expressed in a ratio form as already discussed. Similar to the monthly data regime, the x-axis of Figure 6.5 and Figure 6.6 represent the dates of the forecasting evaluation time period whilst the y-axis indicate the associated value of CMSFE for each specification model in the empirical exercise.

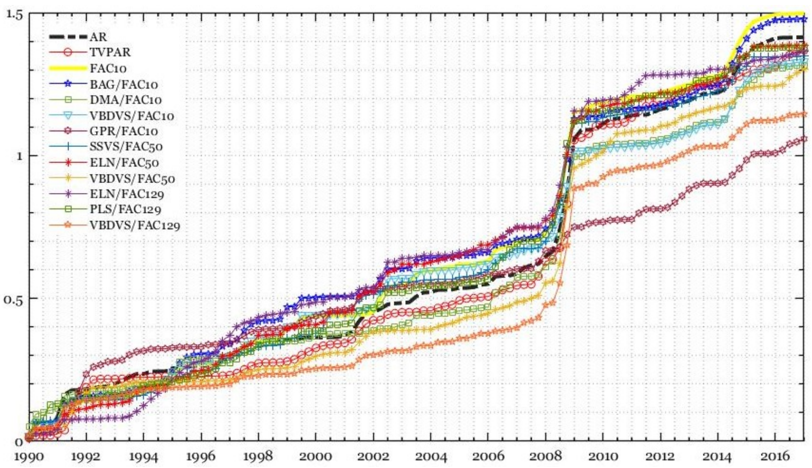


Figure 6.5: Graphical representation of CPI cumulative sum of quarterly squared forecast errors for the evaluation period of $h = 4$.

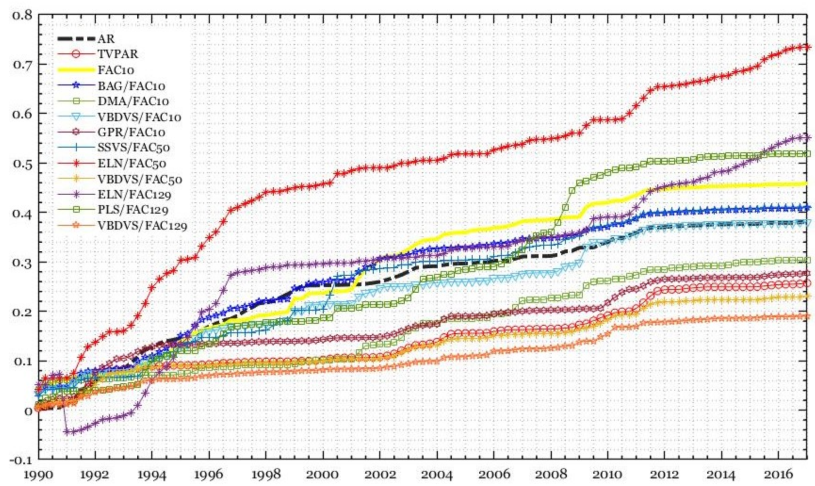


Figure 6.6: Graphical representation of PCEPI cumulative sum of quarterly squared forecast errors for the evaluation period of $h = 4$.

Table 6.9: Quarterly forecasting results CPI inflation measure

	MSFE					ALPL				
	h=1	*h=2*	*h=3*	*h=4*	*h=8*	*h=1*	*h=2*	*h=3*	*h=4*	*h=8*
MODLES WITH ZERO PREDICTORS										
AR	*0.169*	*0.187*	*0.144*	*0.126*	*0.121*	*4.352*	*4.422*	*4.455*	*4.377*	*4.451*
TVPAR	1.041	1.119	1.064	0.970	0.619	0.076	0.337	0.220	0.132	0.574
10 FACTOR MODELS										
FAC10	0.994	1.013	1.055	1.059	1.044	0.070	0.066	0.032	0.079	0.105
BAG/FAC 10	**0.946**	0.984	1.015	1.045	1.143	0.065	0.051	0.038	0.056	-0.008
DMA/FAC10	1.015	0.967	0.964	0.931	0.985	0.110	0.229	0.188	0.204	0.118
GPR/FAC10	1.025	0.980	0.952	0.941	1.058	0.024	0.015	0.003	0.071	0.060
VBDVS/FAC10	0.977	1.057	0.927	0.852	0.696	0.134	0.273	0.229	-0.040	0.196
50 FACTOR MODELS										
SSVS/FAC50	1.043	0.876	**0.902**	0.954	0.875	**0.098**	0.102	0.162	0.210	0.128
ELN/FAC50	1.150	0.870	0.952	0.981	0.865	0.045	0.068	0.149	0.209	0.222
VBDVS/FAC50	0.988	1.078	0.967	0.923	0.594	0.097	**0.276**	0.482	0.320	0.924
129 FACTOR MODELS										
ELN/129	1.045	1.004	0.969	0.967	1.017	0.076	-0.014	0.047	0.102	0.543
PLS/129	1.306	0.982	0.952	0.977	0.926	0.038	0.220	0.362	0.144	-0.513
VBDVS/129	1.054	**0.934**	0.928	**0.812**	**0.585**	0.036	0.186	**0.622**	**0.391**	**0.977**

*Boldfaced entities signify the dominant forecasting model for every forecast horizon and statistic.

Table 6.10: Quarterly forecasting results for PCEPI inflation measure

	MSFE					**ALPL**				
	h=1	*h=2*	*h=3*	*h=4*	*h=8*	*h=1*	*h=2*	*h=3*	*h=4*	*h=8*
MODLES WITH ZERO PREDICTORS										
AR	*0.024*	*0.022*	*0.027*	*0.034*	*0.068*	*4.804*	*4.888*	*4.862*	*4.782*	*4.715*
TVPAR	0.868	0.725	0.773	0.676	0.407	0.361	0.565	0.430	0.723	0.468
10 FACTOR MODELS										
FAC10	1.331	1.360	1.315	1.203	0.804	0.016	0.057	0.084	0.066	0.140
BAG/FAC 10	1.062	1.209	1.129	1.076	0.830	0.059	0.112	0.136	0.112	0.151
DMA/FAC10	1.150	0.992	0.886	0.799	0.647	0.281	0.393	0.406	0.393	0.123
GPR/FAC10	1.037	1.007	0.983	0.999	0.751	0.128	0.210	0.211	0.154	0.115
VBDVS/FAC10	1.359	1.198	0.973	0.727	0.426	0.195	0.500	0.569	0.665	0.322
50 FACTOR MODELS										
SSVS/FAC50	1.123	1.231	1.221	1.080	0.716	0.085	0.110	0.167	0.224	0.318
ELN/FAC50	2.829	2.782	2.449	1.923	1.042	-0.279	-0.246	-0.261	-0.061	0.097
VBDVS/FAC50	**0.874**	0.708	0.714	0.606	0.397	0.443	**0.829**	**0.929**	**0.972**	0.679
129 FACTOR MODELS										
ELN/129	0.968	0.817	1.340	1.442	1.479	-0.321	-0.364	-0.163	0.632	-0.373
PLS/129	2.062	2.131	1.663	1.360	1.010	-0.090	-0.064	0.020	0.063	0.044
VBDVS/129	1.127	**0.685**	**0.597**	**0.498**	**0.359**	**0.530**	0.749	0.827	0.792	**0.898**

*Boldfaced entities signify the dominant forecasting model for every forecast horizon and statistic.

6.2.1 Examining Inflation Forecast Gains For Quarterly Data

Inspecting the performance estimates, the indication from the empirical data tables includes the predominant forecasting performance of VBDVS/129 model, emphasised by the point forecast estimations for longer forecasting horizons $h = 4, 8$. Additionally, the superior density forecasts of both the VBDVS/50 and VBDVS/129 models, jointly classify them as the more efficient forecasting specifications. Additionally, it is crucial to emphasise that overall, models encoding an increased predictor number namely - the 129 factor and 50 factor model regimes, display enhanced forecasting gains specifically for longer time horizons in contrast to the TVP-AR specification including either 10 explanatory variables or no explanatory variables. The underlying finding is also in line with the aforementioned monthly empirical exercise.

One can derive numerous stylised facts from the information included in the empirical findings. Nonetheless, the current discussion emphasises the point forecasts metric as it outlines a more comprehensive picture. According to the empirical findings, the incorporation of time variation appears to substantially increase forecasting performance, especially in longer evaluation times. In particular, TVPAR, DMA/FAC10, and all VBDVS model specifications experience radical improvements compared to their corresponding constant parameter counterparts, irrespective of the predictor classification - exogenous or endogenous. The exact determination of the forecasting advances in regards to exogenous predictors, strongly depends on the particular forecasting objectives namely - the predictor treatment and incorporation process of a particular specification model and the forecasting horizon of interest. Inspecting the empirical results, the substantial difference in the forecasting performance for $h = 8$, between the TVP model and VBDVS model specification, which essentially corresponds to a TVP model with all exogenous predictors incorporated, is obtained in Table 6.9 and Table 6.10 intimating the forecasting advances of both time-variation and information incorporated in exogenous predictors. On the contrary, the forecasting performance of constant parameter models with exogenous predictors, being either observed or included in regression specifications via factor approaches, appears to be inferior to the benchmark AR(2) model.

Furthermore, an increasing number of exogenous parameters does appear to be significant in terms of improving forecasting performance for the initial time horizon forecast evaluations, in regards to long-run forecasting however, the greatest contribution in forecasting accuracy is denoted by inter alia by the incorporation of time variation in the parameter coefficients. For instance, inspecting both the CPI and PCEPI point forecasts for $h = 1, 2$ the ELN/129 model specification outperforms both the AR(2) and TVPAR models. Yet, in the following time horizons $h = 3, 4, 8$, the TVPAR specification significantly surpasses both the ELN/129 and AR(2) models. Furthermore, even though the VBDVS/129 model specification dominates all competing model regimes for longer time horizons, the statistical discrepancies in regards to the TVPAR model performance are less significant than the ones associated with the ELN/129 and AR(2) models. Regardless, of the predictor's forecasting significance, the exceptional shrinkage advances of the VB-DVS algorithm are evident throughout the current empirical application reflected by its forecasts superiority to the TVPAR model. Hence, no overfitting incidents emerge in the examined regime.

Finally, no stylised facts are extracted for inflation forecasting on the basis of the ALPL estimates as the associated metric emanates on the entire predictive density spectrum comprising of both the mean and the dynamics of higher moments. Furthermore, accounting for the wide variation in predictive densities, the ALPLs difference attribution is not computationally feasible. Overall, the standard pattern emerging through the current empirical application is the inflation forecasting advancement resulting from the inclusion of exogenous predictors, contributing to density forecasts improvements relative to the benchmark model. However, the greatest forecasting gains are accomplished via the implementation time variation in exogenous parameters. In terms of the competing model specifications, the VB-DVS/129 model, regardless of the fact that is heavily parametrized, it dominates all other models due to the implementation of an efficient penalisation regime giving rise to exceptional forecasting gains and accordingly to erroneous forecasts estimates in regards to quarterly inflation forecasting.

7 | Conclusion

Overall, the current investigation evaluates and compares a wide range of forecasting specifications via the implementation of an explanatory predictor set composed of 129 real-activity macroeconomic time series ranging form $1960M1-2017M12$. Specifically, two empirical applications were engaged, the first focuses on monthly inflation forecasts while the second employs a quarterly forecasting regime. However, the obtained empirical findings do not outline a single consistent inference but, rather an assortment of inferences as a significant sampling variability in forecast comparison measures is identified in the empirical results.

Recalling the initial monthly forecasting application based on the regression specification denoted by Equation (5.1), the empirical findings regarding the average point forecasts outline the forecasting superiority of approximations entailing a significant number of predictors namely -BMA, TVP-BMA and TVP-GMAP, denoting increasing forecasting gains for longer time horizons. Furthermore, substantial forecasting advantages are obtained in forecasting specifications incorporating a stochastic volatility term which appears to minimise potential model misspecifications. Moreover, approximation regimes that impose restrictions on the variation magnitude of time-varying parameters dominate approaches controlling for overparameterization issues. However, time-varying dimension model specifications displayed a poor performance in the underlying regime suffering from overfitting errors. Emphasis should also attributed to the fact that greater forecasting relevance has found to be associated with the information enclosed in the explanatory predictors and not as much to the evolution approximation of the time-varying parameters.

Additionally, a remarkable finding of the current study includes the improvement of all model specifications using regression Equation (5.2), signifying an improved goodness of fit especially regarding structural-break and time-varying dimension approximations - KP-AR, GK-AR and TVD. These advances are associated with the augmentation of the random walk dynamics in the Phillips curve specification. Specifically, models allowing for inclusion parameter variations are ranked amongst the most suitable candidates for inflation forecasting along with regression averaging models by optimising all risks associated with model uncertainty. An additional key finding is that even though the vast majority of point forecast evaluations appeared to outperform the simple univariate AR(2) benchmark specification, there are times where the benchmark AR(2) outperformed all other forecasting approaches denoting the episodic nature of Phillips curve and confirming the declaration of Atkeson et al. (2001) concerning Phillips curve based forecasts.

Moving on, the extraction of stylized facts in regards to inflation forecasting accordance with the LAPLs is more cumbersome since the examined metric relies on the complete feature spectrum of the predictive density. Nevertheless, the empirical findings associated with density forecast realisations using regression Equation (5.2), denote once again the substandard forecasting performance of models with a small predictor number compared to the benchmark specification. However, once the alternative regression regime is implemented the majority of models' predictive accuracies appear to outperform the AR(2) benchmark with increasing accuracy gains in longer forecast horizons. Yet, the empirical findings do not distinguish a single dominant inflation forecasting specification as models' performances vary over time, but on average the most accurate forecasting model specifications displaying minimum discrepancies in their forecasting abilities include the TVP-GAMP and TVP-BMA models.

The forecasting meaningfulness of the explanatory predictor sample size was further investigated by the second empirical exercise where four different model subsets were investigated. The first incorporated no predictors, the second incorporated 10 predictors, the third incorporated 50 predictors and the last incorporated all 129 explanatory predictors. The associated empirical findings demonstrate that overall high-dimensional forecasting regimes, even though they tend to suffer from heavy parameterization problems, dominate competing approximations with the VBDVS algorithm displaying the greatest forecasting performance. Lastly, it should be emphasised that comparing the two forecasting regimes the cumulative sum of MSFEs associated with quarterly forecast appears to be lower than the cumulative sums of MSFEs associated with the monthly forecasting exercise, suggesting that a quarterly forecasting frequency might be more suitable for inflation forecasting.

Bibliography

Al-Shoukairi, M., Schniter, P. & Rao, B. D. (2017), 'A gamp-based low complexity sparse bayesian learning algorithm', *IEEE Transactions on Signal Processing* **66**(2), 294–308.

Angelino, E., Johnson, M. J. & Adams, R. P. (2016), 'Patterns of scalable bayesian inference', *arXiv preprint arXiv:1602.05221* .

Atkeson, A., Ohanian, L. E. et al. (2001), 'Are phillips curves useful for forecasting inflation?', *Federal Reserve bank of Minneapolis quarterly review* **25**(1), 2–11.

Bauwens, L., Koop, G., Korobilis, D. & Rombouts, J. V. (2015), 'The contribution of structural break models to forecasting macroeconomic series', *Journal of Applied Econometrics* **30**(4), 596–620.

Beal, M. J. (2003), Variational algorithms for approximate Bayesian inference, PhD thesis, UCL (University College London).

Bernanke, B. & Mishkin, F. (1992), 'Central bank behavior and the strategy of monetary policy: observations from six industrialized countries', *NBER Macroeconomics Annual* **7**, 183–228.

Berry, A. C. (1941), 'The accuracy of the gaussian approximation to the sum of independent variates', *Transactions of the american mathematical society* **49**(1), 122–136.

Blanchard, O. (2016), 'The phillips curve: Back to the'60s?', *American Economic Review* **106**(5), 31–34.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994), 'Arch models', *Handbook of econometrics* **4**, 2959–3038.

Box, G., Jenkins, G. & Reinsel, G. (1994), 'Time series analysis, forecasting and control. englewood clifs'.

Breiman, L. (1999), Using adaptive bagging to debias regressions, Technical report, Technical Report 547, Statistics Dept. UCB.

Burns, A. F. & Mitchell, W. C. (1946), 'Measuring business cycles'.

Campbell, J. Y. & Mankiw, N. G. (1987), 'Are output fluctuations transitory?', *The Quarterly Journal of Economics* **102**(4), 857–880.

Chan, J. C., Koop, G., Leon-Gonzalez, R. & Strachan, R. W. (2012), 'Time varying dimension models', *Journal of Business & Economic Statistics* **30**(3), 358–367.

Chan, Y. L., Stock, J. H. & Watson, M. W. (1999), 'A dynamic factor model framework for forecast combination', *Spanish Economic Review* **1**(2), 91–121.

Chib, S. (1998), 'Estimation and comparison of multiple change-point models', *Journal of econometrics* **86**(2), 221–241.

Clark, T. E. & Ravazzolo, F. (2015), 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics* **30**(4), 551–575.

Clements, M. & Hendry, D. (1998), *Forecasting economic time series*, Cambridge University Press.

Cogley, T., Primiceri, G. E. & Sargent, T. J. (2010), 'Inflation-gap persistence in the us', *American Economic Journal: Macroeconomics* **2**(1), 43–69.

Cogley, T. & Sargent, T. J. (2001), 'Evolving post-world war ii us inflation dynamics', *NBER macroeconomics annual* **16**, 331–373.

Cogley, T. & Sargent, T. J. (2005), 'Drifts and volatilities: monetary policies and outcomes in the post wwii us', *Review of Economic dynamics* **8**(2), 262–302.

Diebold, F. M. & Mariano, R. (1995), 'R.(1995). comparing predictive accuracy', *Journal of Business & economic statistics* **20**(1).

Diebold, F. X. (1998), 'The past, present, and future of macroeconomic forecasting', *Journal of Economic Perspectives* **12**(2), 175–192.

Diebold, F. X. & Rudebusch, G. D. (1994), Measuring business cycles: A modern perspective, Technical report, National Bureau of Economic Research.

Diebold, F. X. & Senhadji, A. S. (1996), 'The Uncertain Unit Root in Real GNP: Comment', *American Economic Review* **86**(5), 1291–1298.
URL: *https://ideas.repec.org/a/aea/aecrev/v86y1996i5p1291-98.html*

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Gelman, A., Vehtari, A., Jylänki, P., Robert, C., Chopin, N. & Cunningham, J. P. (2014), 'Expectation propagation as a way of life', *arXiv preprint arXiv:1412.4869* **157**.

George, E. I. & McCulloch, R. E. (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **88**(423), 881–889.

Gerlach, R., Carter, C. & Kohn, R. (2000), 'Efficient bayesian inference for dynamic mixture models', *Journal of the American Statistical Association* **95**(451), 819–828.

Geweke, J. & Amisano, G. (2010), 'Comparing and evaluating bayesian predictive distributions of asset returns', *International Journal of Forecasting* **26**(2), 216–230.

Giordani, P. & Kohn, R. (2008), 'Efficient bayesian inference for multiple changepoint and mixture innovation models', *Journal of Business & Economic Statistics* **26**(1), 66–77.

Gordon, R. J. (1997), 'The time-varying nairu and its implications for economic policy', *Journal of economic Perspectives* **11**(1), 11–32.

Gordon, R. J. & Stock, J. H. (1998), 'Foundations of the goldilocks economy: supply shocks and the time-varying nairu', *Brookings papers on economic activity* **1998**(2), 297–346.

Granger, C. W. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica: journal of the Econometric Society* pp. 424–438.

Granger, C. W., Terasvirta, T. et al. (1993), 'Modelling non-linear economic relationships', *OUP Catalogue* .

Groen, J. J., Paap, R. & Ravazzolo, F. (2013), 'Real-time inflation forecasting in a changing world', *Journal of Business & Economic Statistics* **31**(1), 29–44.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica: Journal of the Econometric Society* pp. 357–384.

Heij, C., Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K. et al. (2004), *Econometric methods with applications in business and economics*, Oxford University Press.

Hershey, J. R. & Olsen, P. A. (2007), Approximating the kullback leibler divergence between gaussian mixture models, *in* '2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07', Vol. 4, IEEE, pp. IV–317.

Inoue, A. & Kilian, L. (2008), 'How useful is bagging in forecasting economic time series? a case study of us consumer price inflation', *Journal of the American Statistical Association* **103**(482), 511–522.

Kalli, M. & Griffin, J. E. (2014), 'Time-varying sparsity in dynamic regression models', *Journal of Econometrics* **178**(2), 779–793.

Kamilov, U., Goyal, V. K. & Rangan, S. (2011), Optimal quantization for compressive sensing under message passing reconstruction, *in* '2011 IEEE International Symposium on Information Theory Proceedings', IEEE, pp. 459–463.

Karush, J. (1961), 'On the chapman-kolmogorov equation', *The Annals of Mathematical Statistics* **32**(4), 1333–1337.

Kenessey, Z. (1987), 'The primary, secondary, tertiary and quaternary sectors of the economy', *Review of Income and Wealth* **33**(4), 359–385.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: likelihood inference and comparison with arch models', *The review of economic studies* **65**(3), 361–393.

Klein, L. R. & Goldberger, A. S. (1955), 'Econometric model of the united states, 1929-1952'.

Koop, G. & Korobilis, D. (2010), *Bayesian multivariate time series methods for empirical macroeconomics*, Now Publishers Inc.

Koop, G. & Korobilis, D. (2012), 'Forecasting inflation using dynamic model averaging', *International Economic Review* **53**(3), 867–886.

Koop, G. & Potter, S. M. (2007), 'Estimation and forecasting in models with multiple breaks', *The Review of Economic Studies* **74**(3), 763–789.

Korobilis, D. (2013), 'Hierarchical shrinkage priors for dynamic regressions with many predictors', *International Journal of Forecasting* **29**(1), 43–59.

Korobilis, D. (2019), 'High-dimensional macroeconomic forecasting using message passing algorithms', *Journal of Business & Economic Statistics* pp. 1–12.

Korobilis, D. & Koop, G. (2020), 'Bayesian dynamic variable selection in high dimensions'.

Kschischang, F. R., Frey, B. J. & Loeliger, H.-A. (2001), 'Factor graphs and the sum-product algorithm', *IEEE Transactions on information theory* **47**(2), 498–519.

Kulhavỳ, R. & Kraus, F. J. (1996), 'On duality of regularized exponential and linear forgetting', *Automatica* **32**(10), 1403–1415.

Leamer, E. E. & Leamer, E. E. (1978), *Specification searches: Ad hoc inference with nonexperimental data*, Vol. 53, Wiley New York.

Lucas, R. E. et al. (1976), Econometric policy evaluation: A critique, *in* 'Carnegie-Rochester conference series on public policy', Vol. 1, pp. 19–46.

McConnell, M. M. & Perez-Quiros, G. (2000), 'Output fluctuations in the united states: What has changed since the early 1980's?', *American Economic Review* **90**(5), 1464–1476.

Mitchell, W. C. & Burns, A. F. (1938), Statistical indicators of cyclical revivals, *in* 'Statistical indicators of cyclical revivals', NBER, pp. 1–12.

Mooij, J. M. & Kappen, H. J. (2007), 'Sufficient conditions for convergence of the sum–product algorithm', *IEEE Transactions on Information Theory* **53**(12), 4422–4437.

Narisetty, N. N., He, X. et al. (2014), 'Bayesian variable selection with shrinking and diffusing priors', *The Annals of Statistics* **42**(2), 789–817.

Nelson, C. R. & Plosser, C. R. (1982), 'Trends and random walks in macroeconmic time series: some evidence and implications', *Journal of monetary economics* **10**(2), 139–162.

Pearl, J. (1982), *Reverend Bayes on inference engines: A distributed hierarchical approach*, Cognitive Systems Laboratory, School of Engineering and Applied Science â¦.

Pesaran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), 'Forecasting time series subject to multiple structural breaks', *The Review of Economic Studies* **73**(4), 1057–1084.

Pettenuzzo, D. & Timmermann, A. (2017), 'Forecasting macroeconomic variables under model instability', *Journal of Business & Economic Statistics* **35**(2), 183–201.

Pitt, M. K., Chatfield, C. & Walker, S. G. (2002), 'Constructing first order stationary autoregressive models via latent processes', *Scandinavian Journal of Statistics* **29**(4), 657–663.

Pitt, M. K. & Walker, S. G. (2005), 'Constructing stationary time series models using auxiliary variables with applications', *Journal of the American Statistical Association* **100**(470), 554–564.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Quah, D. (1993), 'Empirical cross-section dynamics in economic growth', *European Economic Review* **37**(2-3), 426–434.
URL: *https://EconPapers.repec.org/RePEc:eee:eecrev:v:37:y:1993:i:2-3:p:426-434*

Raftery, A. E., Kárnỳ, M. & Ettler, P. (2010), 'Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill', *Technometrics* **52**(1), 52–66.

Rangan, S. (2011), Generalized approximate message passing for estimation with random linear mixing, *in* '2011 IEEE International Symposium on Information Theory Proceedings', IEEE, pp. 2168–2172.

Rangan, S., Schniter, P., Riegler, E., Fletcher, A. K. & Cevher, V. (2016), 'Fixed points of generalized approximate message passing with arbitrary matrices', *IEEE Transactions on Information Theory* **62**(12), 7464–7474.

Roberts, J. M. (2004), 'Monetary policy and inflation dynamics'.

Rockova, V., McAlinn, K. et al. (2020), 'Dynamic variable selection with spike-and-slab process priors', *Bayesian Analysis* .

Sargent, T. J., Sims, C. A. et al. (1977), 'Business cycle modeling without pretending to have too much a priori economic theory', *New methods in business cycle research* **1**, 145–168.

Sargent, T. & Sims, C. (1977), Business cycle modeling without pretending to have too much a priori economic theory, Working Papers 55, Federal Reserve Bank of Minneapolis.
URL: *https://EconPapers.repec.org/RePEc:fip:fedmwp:55*

Sims, C. A. (1972), 'Money, income, and causality', *The American economic review* **62**(4), 540–552.

Sims, C. A. & Zha, T. (2006), 'Were there regime switches in us monetary policy?', *American Economic Review* **96**(1), 54–81.

Slutzky, E. (1937), 'The summation of random causes as the source of cyclic processes', *Econometrica: Journal of the Econometric Society* pp. 105–146.

Šmídl, V. & Quinn, A. (2006), *The variational Bayes method in signal processing*, Springer Science & Business Media.

Staiger, D. O., Stock, J. H. & Watson, M. W. (1997), How precise are estimates of the natural rate of unemployment?, *in* 'Reducing inflation: Motivation and strategy', University of Chicago Press, pp. 195–246.

Stock, J. H. & Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *NBER macroeconomics annual* **4**, 351–394.

Stock, J. H. & Watson, M. W. (1996), 'Evidence on structural instability in macroeconomic time series relations', *Journal of Business & Economic Statistics* **14**(1), 11–30.

Stock, J. H. & Watson, M. W. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**(2), 293–335.

Stock, J. H. & Watson, M. W. (2007), 'Why has us inflation become harder to forecast?', *Journal of Money, Credit and banking* **39**, 3–33.

Stock, J. H. & Watson, M. W. (2008), Phillips curve inflation forecasts, Technical report, National Bureau of Economic Research.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Tipping, M. E. (2001), 'Sparse bayesian learning and the relevance vector machine', *Journal of machine learning research* **1**(Jun), 211–244.

Tong, H. (1990), *Non-linear time series: a dynamical system approach*, Oxford University Press.

Wand, M. P. (2017), 'Fast approximate inference for arbitrarily large semiparametric regression models via message passing', *Journal of the American Statistical Association* **112**(517), 137–168.

Wang, H., Yu, H., Hoy, M., Dauwels, J. & Wang, H. (2016), Variational bayesian dynamic compressive sensing, *in* '2016 IEEE International Symposium on Information Theory (ISIT)', IEEE, pp. 1421–1425.

West, M. & Harrison, J. (2006), *Bayesian forecasting and dynamic models*, Springer Science & Business Media.

Wright, J. H. (2009), 'Forecasting us inflation by bayesian model averaging', *Journal of Forecasting* **28**(2), 131–144.

Yule, G. U. (1927), 'Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers', *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **226**(636-646), 267–298.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the royal statistical society: series B (statistical methodology)* **67**(2), 301–320.

Zou, X., Li, F., Fang, J. & Li, H. (2016), Computationally efficient sparse bayesian learning via generalized approximate message passing, *in* '2016 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)', IEEE, pp. 1–4.

A | Data-set

Table A.1: Macroeconomic data set based on monthly US data extracted form FRED.

No	Mnemonic	Long Description	Tcode
		Exchange Rates and Interest Rates	
1	FEDFUNDS	Effective Federal Funds Rate	2
2	AAA	Aaa Corporate Bond Yield	2
3	AAAFFM	Aaa - FFR spread	1
4	BAA	Baa Corporate Bond Yield	2
5	BAAFFM	Baa - FFR spread	1
6	COMPAPFFx	CP - FFR spread	1
7	CP3Mx	3-Month AA Comm. Paper Rate	2
8	EXCAUSx	Canada / U.S. FX Rate	5
9	EXJPUSx	Japan / U.S. FX Rate	5
10	EXSZUSx	Switzerland / U.S. FX Rate	5
11	EXUSUKx	U.S. / U.K. FX Rate	5
12	GS1	1-Year T-bond	2
13	GS10	10-Year T-bond	2
14	GS5	5-Year T-bond	2
15	T10YFFM	10 yr. - FFR spread	1
16	T1YFFM	1 yr. - FFR spread	1
17	T5YFFM	5 yr. - FFR spread	1
18	TB3MS	3-Month T-bill	2
19	TB3SMFFM	3 Mo. - FFR spread	1
20	TB6MS	6-Month T-bill	2
21	TB6SMFFM	6 Mo. - FFR spread	1

Continued on next page

Table A.1 – continued from previous page

No	Mnemonic	Long Description	Tcode
		Housing Starts and Permits	
22	HOUST	Starts: Total	4
23	HOUSTMW	Starts: Midwest	4
24	HOUSTNE	Starts: Northeast	4
25	HOUSTS	Starts: South	4
26	HOUSTW	Starts: West	4
27	PERMIT	Permits	4
28	PERMITMW	Permits: Midwest	4
29	PERMITNE	Permits: Northeast	4
30	PERMITS	Permits: South	4
31	PERMITW	Permits: West	4
		Income and Output	
32	CAPUTLB00004S	Capacity Utilisation:Manufacturing	1
33	INDPRO	IP Index	5
34	IPB51222S	IP: Residential Utilities	5
35	IPBUSEQ	IP: Business Equipment	5
36	IPCONGD	IP: Consumer Goods	5
37	IPDCONGD	IP: Durable Consumer Goods	5
38	IPDMAT	IP: Durable Materials	5
39	IPFINAL	IP: Final Products	5
40	IPFPNSS	IP: Final Products and Supplies	5
41	IPFUELS	IP: Fuels	5
42	IPMANSICS	IP: Manufacturing	5
43	IPMAT	IP: Materials	5
44	IPNCONGD	IP: Nondurable Consumer Goods	5

Continued on next page

Table A.1 – continued from previous page

No	Mnemonic	Long Description	Tcode
45	IPNMAT	IP: Nondurable Materials	5
46	NAPMPI	ISM Manufacturing: Production	1
47	RPI	Real Personal Income	5
48	W875RX1	RPI ex. Transfers	5
		Labor Market	
49	HWI	Help-Wanted Index for US	1
50	AWHMAN	Hours: Manufacturing	1
51	AWOTMAN	Overtime Hours: Manufacturing	1
52	CE16OV	Civilian Employment	5
53	CES06000007	Hours: Goods-Producing	1
54	CES06000008	Average Hourly Earnings: Goods	5
55	CES102100001	All Employees: Mining and Logging	6
56	CES20000008	Average Hourly Earnings: Construction	6
57	CES20000008	Average Hourly Earnings: Manufacturing	6
58	CLAIMSx	Initial Claims	5
59	CLF16OV	Civilian Labor Force	5
60	DMANEMP	All Employees: Durable goods	5
61	HWIURATIO	Help Wanted to Unemployed ratio	1
62	MANEMP	All Employees: Manufacturing	5
63	NAPMEI	ISM Manufacturing: Employment	1
64	NDMANEMP	All Employees: Nondurable goods	5
65	PAYEMS	All Employees: Total nonfarm	5
66	SRVPRD	All Employees: Service Industries	5
67	UEMP15OV	Civilians Unemployed > 15 Weeks	5
68	UEMP15T26	Civilians Unemployed 15-26 Weeks	5
69	UEMP27OV	Civilians Unemployed > 27 Week	5

Continued on next page

Table A.1 – continued from previous page

No	Mnemonic	Long Description	Tcode
70	UEMP5TO14	Civilians Unemployed 5-14 Weeks	5
71	UEMPLT5	Civilians Unemployed $\leq$ 5 Weeks	5
72	UEMPMEAN	Average Duration of Unemployment	1
73	UNRATE	Civilian Unemployment Rate	1
74	USCONS	All Employees: Construction	5
75	USFIRE	All Employees: Financial Activities	5
76	USGOOD	All Employees: Goods-Producing	5
77	USGOVT	All Employees: Government	5
78	USTPU	All Employees: TT&U	5
79	USTRADE	All Employees: Retail Trade	5
80	USWTRADE	All Employees: Wholesale Trade	5
		Money and Credit	
81	M1SL	M1 Money Stock	5
82	AMBSL	St. Louis Adjusted Monetary Base	5
83	BUSLOANS	Industrial and Commercial Loans	5
84	CONSPI	Credit to PI ratio	1
85	DTCOLNVHF	Consumer Motor Vehicle Loans	6
86	DTCTHFNM	Total Consumer Loans and Leases	6
87	INVEST	Securities in Bank Credit	6
88	M2REAL	Real M2 Money Stock	5
89	M2SL	M2 Money Stock	5
90	MZMSL	MZM Money Stock	6
91	NONREVSL	Total Nonrevolving Credit	5
92	REALLN	Real Estate Loans	5
93	TOTRESNS	Total Reserves	5

Continued on next page

Table A.1 – continued from previous page

No	Mnemonic	Long Description	Tcode
		Prices	
94	CPIAPPSL	CPI: Apparel	6
95	CPIAUCSL	CPI: All Items	6
96	CPIMEDSL	CPI: Medical Care	6
97	CPITRNSL	CPI: Transportation	6
98	CPIULFSL	CPI: All Items Less Food	6
99	CUSR0000SA0L5	CPI: All items less medical care	6
100	CUSR0000SAC	CPI: Commodities	6
101	CUSR0000SAS	CPI: Services	6
102	CUUR0000SA0L2	CPI: All items less shelter	6
103	CUUR0000SAD	CPI: Durables	6
104	DDURRG3M086S	PCE: Durable goods	6
105	DNDGRG3M086S	PCE: Nondurable goods	6
106	DSERRG3M086S	PCE: Services	6
107	PCEPI	PCE: Chain-type Price Index	6
108	PPICMM	PPI: Commodities	6
109	PPICRM	PPI: Crude Materials	6
110	PPIFCG	PPI: Finished Consumer Good	6
111	PPIFGS	PPI: Finished Goods	6
112	PPIITM	PPI: Intermediate Materials	6
113	NAPMPRI	ISM Manufacturing: Prices	1
114	OILPRICEx	Crude Oil Prices: WTI	6
		Stock Market	
115	S&P 500	S&P: Composite	5
116	S&P div yield	S&P: Dividend Yield	1

Continued on next page

Table A.1 – continued from previous page

No	Mnemonic	Long Description	Tcode
117	S&P indust	S&P: Industrials	5
118	S&P PE ratio	S&P: Price-Earnings Ratio	5
		Orders, Consumption and Inventories	
119	DPCERA3M086S	Real PCE	5
120	AMDMNOx	Orders: Durable Goods	5
121	AMDMUOx	Unfilled Orders: Durable Goods	5
122	BUSINVx	Total Business Inventories	5
123	CMRMTSPLx	Real M&T Sale	5
124	ISRATIOx	Inventories to Sales Ratio	1
125	NAPMII	ISM: Inventories Index	1
126	NAPMNOI	ISM: New Orders Index	1
127	NAPMPI	ISM: PMI Composite Index	1
128	NAPMSDI	ISM: Supplier Deliveries Index	1
129	RETAILx	Retail and Food Services Sales	5

B | Analytical Derivation of GAMP

At the core of the examined algorithm lies the fundamental signal extraction problem often observed in the field of engineering. Adopting the stereotypical regression regime, the problem requires the input of observation data y along with a determined "transformation-matrix" x and a signal β. Additionally, the assumption of an "adaptive white Gaussian noise" (AWGN) is implemented in the examination regime denoting the disturbance term encountered in signal processing applications. Hence, the assumed model is given by:

$$y = x\beta + \varepsilon \ , \tag{B.1}$$

where y and β denote a $T \times 1$ and $q \times 1$ vectors respectively, x and β is a Tq matrix and $\varepsilon \sim N(0, \sigma \times I_T)$ of size $T \times 1$. Recalling the TVP regression Equation (4.10) by replacing x with $\mathscr{X}$ the problem can be represented by the general form equation. Moreover, the likelihood function is defined as $p(y|z)$ where $z = x\beta$. As already mentioned the construction of the algorithm is build upon the factor graph methodology. Assuming a message function propagating from node a to node b $\mu_{a\to b}$ whose logarithmic form is denoted by $m_{a\longrightarrow b}$ the initial algorithm estimation includes the counter-cyclical sum-product Belief Propagation whose logarithmic form is expressed as:

$$m^{(r+1)}_{p(y_t|z_t)\longrightarrow\beta_i} = \log\left(\int p(y_t|z_t) \prod_{j=1,j\neq i}^{q} \mu_{\beta_j\to p(y_t|z_t)} d\beta_{j\neq i}\right) = \log E_{z_t}[p(y_t, z_t)] \ , \tag{B.2}$$

$$m^{(r+1)}_{\beta_i\longrightarrow p(y_t|z_t)} = \log(p(\beta_i)) + \sum_{s=1,s\neq t}^{T} m^{(r)}_{p(y_s|\beta)\to\beta_i} \ , \tag{B.3}$$

where the expectation of $p(y_t|z_t)$ with regards to β is denoted by $E_{z_t}[p(y_t|z_t)]$. Thus, the approximation of the marginal posterior under this regime is expressed by:

$$p(\beta_i|y) \propto exp(m_i) \quad , \tag{B.4}$$

where the expression form m_i is given by:

$$m_i \equiv \log(p(\beta_i)) + \sum_{t=1}^{T} m_{p(y_t|z_t)\rightarrow\beta_i} \; . \tag{B.5}$$

There is a substantial variety of algorithms serving the approximation purpose of the sum-product iterations including the Mean Field Variational Bayes (MFVB) class suitable for hierarchical regression models. Th key purpose of the GAMP includes the implementation of Gaussian approximations to the Belief Propagation (BP) iterations. All proofs required are based upon the results originally replicated by Rangan (2011) stating form:

Lemma 1. *Employing a random variable U whose conditional probability density function is expressed by:*

$$p(u|v) \equiv \frac{1}{Z(v)} exp(\phi(0) + uv) \quad , \tag{B.6}$$

where, $Z(v)$ is known as the partition function such that:

$$\frac{\partial}{\partial v} lnZ(v) = E(U|V = v) \quad , \tag{B.7}$$

$$\frac{\partial^2}{\partial^2 v} lnZ(v) = \frac{\partial}{\partial v} E(U|V = v) \tag{B.8}$$

$$= var(U|V = v) \tag{B.9}$$

Proof: The above expressions resemble the standard characteristics of exponential families.

Furthermore, considering the following relations:

$$\hat{\beta}_i \equiv E(\beta_i|m_i) \ , \tag{B.10}$$

$$\hat{\tau_i}^\beta \equiv var(\beta_i|m_i) \ , \tag{B.11}$$

$$\hat{\beta}_i \equiv E(\beta_i|m_{\beta_j \to p(y_t|z_t)}) \ , \tag{B.12}$$

$$\hat{\tau}^\beta_{i\to t} \equiv var(\beta_i|m_{\beta_j \to p(y_t|z_t)}) \ , \tag{B.13}$$

where $E(x|f)$ signifies the conditional expectation of random variable x in regards to the density function f.

B.1 Signal messages form factor to variable nodes

The emitted message form the system's "output" nodes can be interpreted in a similar manner as the logarithmic expectation of $p(y_t|z_t)$ with regards to z_t being approximated by $\prod_{j=1,j}^{q} \mu_{\beta_j \to p(y_t|z_t)}$. Considering that the regression likelihood is estimated by:

$$\begin{aligned} m_{p(y_t|z_t)\to\beta_i} &\propto \log E_{zt}[p(y_t|z_t)] && \text{(B.14)}\\ &= \log E_{zt}\left(-\frac{1}{2}\frac{(y_t - z_t)^2}{\sigma^2}\right) && \text{(B.15)}\\ &= \log E_{zt}\left(-\frac{1}{2}\frac{(y_t - x_{t,i}\beta_i - \sum_{j\neq i} x_{t,j}\beta_j)^2}{\sigma^2}\right) \quad , && \text{(B.16)} \end{aligned}$$

where $x_{t,i}\beta_i + \sum_{j\neq i} x_{t,j}\beta_j = \sum_{j=1}^{q} x_{t,j}\beta_j = x_t\beta = z_t$. Additionally, according to the Barry- Esseen central limit theorem, a Normal distribution is reasonably employed for the approximation of, z_t conditional on β_i with the two first moments being estimated as:

$$E(z_t|\beta_i) = x_{t,i}\beta_i + \sum_{j\neq i} x_{t,j} E(\beta_i | m_{\beta_j \to p(y_t|z_t)}) \ , \tag{B.17}$$

$$var(z_t|\beta_i) = x_{t,i}^2 var(\beta_i) + \sum_{j\neq i} x_{t,j}^2 var(\beta_i | m_{\beta_j \to p(y_t|z_t)}) \ . \tag{B.18}$$

Since z_t is conditional on β_i, the variance equation equals zero. Recalling Equations (B.12) and (B.13), the conditional distribution of z_t is expressed by:

$$z_t|\beta_i \sim N\left(x_{t,i}\beta_i + \sum_{j\neq i} x_{t,j}\hat{\beta}_{i\to t},\ \sum_{j\neq i} x_{t,j}^2 \hat{\tau}_{i\to t}^{\beta}\right) \ . \tag{B.19}$$

Hence, the final expression for the BP output message is expressed as follows:

$$m_{p(y_t|z_t)\to\beta_i} \propto \log E_{zt}[p(y_t|z_t)] \tag{B.20}$$

$$\approx \log \int p(y_t|z_t) N\left(x_{t,i}\beta_i + \sum_{j\neq i} x_{t,j}\hat{\beta}_{j\to t}, \sum_{j\neq i} x_{t,j}^2 \hat{\tau}_{j\to t}^{\beta}\right) dz_t \ . \tag{B.21}$$

Moving on, the following functions are defined as:

$$\hat{c}_{t\to i} = \sum_{j\neq i} x_{t,j}\hat{\beta}_{j\to t}, \qquad \hat{\tau}_{t\to i}^{c} = \sum_{j\neq i} x_{t,j}^2 \hat{\tau}_{j\to t}^{\beta}, \tag{B.22}$$

$$\hat{c}_t = \sum_{i=1}^{q} x_{t,i}\hat{\beta}_{i\to t}, \qquad \hat{\tau}_t^{c} = \sum_{i=1}^{q} x_{t,i}^2 \hat{\tau}_{i\to t}^{\beta}, \tag{B.23}$$

$$H\left(\hat{c}, \hat{\tau}^{(c)}, y\right) \equiv \log E(p(y|z)), \qquad g_{out}\left(\hat{c}, \hat{\tau}^{(c)}, y\right) = \frac{\partial}{\partial \hat{c}} H\left(\hat{c}, \hat{\tau}^{(c)}, y\right), \tag{B.24}$$

$$\hat{s}_t = g_{out}\left(\hat{c}, \hat{\tau}_t^{(c)}, y\right), \qquad \tau_t^{s} = \frac{\partial}{\partial \hat{c}} g_{out}\left(\hat{c}, \hat{\tau}_t^{(c)}, y\right) \ . \tag{B.25}$$

Substituting the above definitions into the output signal expression, one arrives to the following result:

$$m_{p(y_t|z_t)\to\beta_i} = H(\hat{c}_{t\to i} + x_{i,t}\beta_i, \hat{\tau}, y) \tag{B.26}$$

$$= H(\hat{c}_t + x_{i,t}(\beta_i - \hat{\beta}_i), \hat{\tau}, y). \tag{B.27}$$

Taking the second order approximation of the above equation one obtains:

$$m_{p(y_t|z_t)\to\beta_i} \approx \hat{\sigma}_t x_{i,t(\beta_i - \hat{\beta}_i)} - \frac{1}{2}\hat{\tau}^s (x_{i,t}(\beta_i - \hat{\beta}_i))^2 \,. \tag{B.28}$$

Hence, according to the above calculation the ingredients required for the estimation of $m_{p(y_t|z_t)\to\beta_i}$ include the estimation of both $\hat{s}$ and $\hat{\tau}^s$ which correspondingly imply the identification of g_{out} function along with its first order derivative. Following the algebraic derivation of Rangan (2011), the obtain expressions include:

$$g_{out}(\hat{c}, \hat{\tau}^c, y) = \frac{1}{\tau^c}(\hat{z} - \hat{c}) \ , \tag{B.29}$$

$$\frac{\partial}{\partial \hat{c}} g_{out}\left(\hat{c}, \hat{\tau}^c_t, y\right) = \frac{1}{\hat{\tau}^c}\left(\frac{1}{\hat{\tau}^c}\hat{\tau}^z - 1\right) \ , \tag{B.30}$$

where $\hat{\tau}^z = var(z|\hat{c}, \hat{\tau}^c) \equiv \frac{\partial}{\partial \hat{c}}\hat{z}$ and $\hat{z} = E(z|\hat{c}, \hat{\tau}^c) \equiv \frac{\int zp(y|z)N(z|\hat{c},\hat{\tau}^c)dz}{\int p(y|z)N(z|\hat{c},\hat{\tau}^c)dz}$.

B.2 Signal messages form variable to factor nodes

Considering now the messages propagating form $m^{(r+1)}_{\beta_j \longrightarrow p(y_t|z_t)}$, recalling Equation (B.3) and substituting the second order Taylor expansion obtained in Equation (B.28), one obtains:

$$m_{\beta_j \longrightarrow p(y_t|z_t)} = \log(p(\beta_i)) + \sum_{s=1, s\neq t}^{T} m_{p(y_s|\beta)\to\beta_i} \tag{B.31}$$

$$\approx \log(p(\beta_i)) + \sum_{s=1, s\neq t}^{T} \left[\hat{s}_s x_{i,s}(\beta_i - \hat{\beta}_i) - \frac{1}{2}\hat{\tau}^s_s (x_{i,s}(\beta_i - \hat{\beta}_i))^2\right]. \tag{B.32}$$

Defining the next set of functions:

$$\hat{d}_{i\to t} = \hat{\tau}^d_{i\to t}\sum_{s=1,s\neq t}^{T}(\hat{s}_s x_{i,s} + \hat{\tau}^s_s x^2_{i,s}\hat{\beta}_i), \quad \hat{\tau}^d_{i\to t} = \left[\sum_{s=1,s\neq t}^{T}\hat{\tau}^s_s x^2_{i,s}\right]^{-1}, \tag{B.33}$$

$$\hat{d} = \hat{\beta}_i + \hat{\tau}^d_i\sum_{t=1}^{T}x_{i,t}\hat{s}_t, \qquad \hat{\tau}^d = \left[\sum_{t=1}^{T}\hat{\tau}^S_t x^2_{i,t}\right]^{-1}, \tag{B.34}$$

$$\hat{d}_{i\to t} = \hat{d}_i - \hat{\tau}^d_i x_{i,t}\hat{s}_t, \qquad \hat{\tau}^d_{i\to t} \approx \hat{\tau}^d_i. \tag{B.35}$$

Therefore, Equation (B.32) can be rearranged to:

$$m_{\beta_j \longrightarrow p(y_t|z_t)} = \log(p(\beta_i)) - \frac{1}{2\hat{\tau}^d_{i\to t}}(\hat{d}_{i\to t-\beta_i})^2 \tag{B.36}$$

$$= \log(p(\beta_i)) - \frac{1}{2\hat{\tau}^d_i}(\hat{d}_i - \hat{\tau}^d_i x_{i,t}\hat{s}_t - \beta_i)^2 \ . \tag{B.37}$$

Finally, the functional forms of the g_{in} and its corresponding derivatives are expressed as:

$$g_{in}(\hat{d}_i, \hat{\tau}^d_i) = E(\beta_i|\hat{d}_i, \hat{\tau}^d_i) \ , \tag{B.38}$$

$$\frac{\partial}{\partial \hat{d}_i}(\hat{d}_i, \hat{\tau}^d_i) = \frac{1}{\hat{\tau}^d_i}var(\beta_i|\hat{d}_i, \hat{\tau}^d_i) \, . \tag{B.39}$$

Following the generic derivation of GAMP, the determination of the exact functional form strongly depends on the examined prior form.

B.3 GAMP implementation using time varying parameter regression combined with a sparse Bayesian learning prior

Employing the following regression model:

$$y_t = x_t\beta + \varepsilon_t \ , \tag{B.40}$$

where y_t vector dimension is equal to $1 \times q$, $\varepsilon_t \sim N(0, \sigma^2_t)$ and the distribution of the prior is assumed to be $\beta \sim N_p(0, \underline{V})$. In Section(4), the predictors matrix was denoted by $\mathscr{X}$ and was constructed with a particular block-diagonal structure. However, the following algorithm is compatible with any non-block or non-sparse diagonal matrix, with the entailed prior being expressed by:

$$p(\beta_i|\alpha_i) \sim N(0, \alpha_i) \ , \tag{B.41}$$

$$p(\alpha_i^{-1}) \sim Gamma(\underline{a}, \underline{b}) \ . \tag{B.42}$$

Additionally, according to the findings of Zou et al. (2016), the accommodation of EM updates of the hyperparameter α_i under this regime can be achieved by augmenting the GAMP algorithm outlined. The optimisation objective subject to α in iteration $(r+1)$ equals to the maximisation of the $Q-$ function expressed by:

$$\alpha^{(r+1)} = \arg\max_{\alpha} Q(\alpha|\alpha^{(r)}) \equiv E_{\alpha}[\log p(\alpha|y, \beta^{(r)})] \ . \tag{B.43}$$

Finally, optimising the above equation with respect to α, by setting the first order derivative equal to zero the obtained condition is expressed as follows:

$$\alpha_i^{(r+1)} = \frac{2\underline{\alpha} - 1}{b\underline{\beta} + \left(\hat{\beta}_i^{(r)}\right)^2} \ , \tag{B.44}$$

where $\hat{\beta}_i^{(r)}$ denotes the estimator of β_i in regards to iteration r, and the derived formula is also found in the corresponding variational regime of Gibbs-sampler updates. The examined algorithm provides an extension to the original message passing algorithm initially proposed by Rangan et al. (2016) incorporating an additional step for the estimation of the stochastic volatility term in combination with an EM-updating step for the determination of Bayesian prior hyperparameter. Finally, as already pre-mentioned the seven mixture approximation components values assigned for parameters μ_i, V_i and π_i based on a single degree of freedom $log\chi^2$ distribution are presented in the Table B.1.

Table B.1: Seven-component mixture approximation based on a single degree of freedom $log\chi^2$ distribution.

Component No.	π_i	μ_i	V_i
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.1795
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.2575	-1.08819	1.26261

C | Dynamic Model Averaging - DMA, Koop and Korobilis 2012

Developing further the DMA methodology initially introduced by Raftery et al. (2010), Koop & Korobilis (2012), introduce a state-space regime involving multiple parallel time-varying parameter regressions denoted by:

$$y_{t+h} = x_t^{(k)}\beta_t^{(k)} + \varepsilon_{t+h} \ , \tag{C.1}$$

$$\beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_t \ , \tag{C.2}$$

where index (k) denotes the corresponding model implemented. The total number of models included in the DMA regime is equal to $K = 2^p$, where p signifies the sub-group all available explanatory variables. The forecast estimation process includes the dynamic averaging over all potential 2^p models. Hence, according to the examined regime index $k = 1, 2, .., K$ denotes every individual TVP regression with a unique predictor number. Accounting for all different coefficients $\beta_t^{(k)}$ and predictor entailed in each individual model, the corresponding variances will inevitably differ such that $var(\eta_t) = Q^{(k)}$ and $var(\varepsilon_{t_h}) = (\sigma_t^2)^{(k)}$. Accordingly, for the efficient evaluation of every potential model combination the current study motivates the implementation of a forgetting factor Kalman filter (FFKF) originally developed by Kulhavỳ & Kraus (1996). Following the DMA literature, after the completion of the estimation process, measures of fit are obtained for every system model at every time period based on the recursive time approximations of both the predictive likelihood and data likelihood via the establishment of a static Bayesian model averaging process, that recursively modifies the predictors entailed in the TVP regression. Additionally, the evolution rates of both the volatility parameters and coefficients are determined by an assortment of decay and forgetting factors including κ and λ. Thus the established default values are in line with the study of Koop & Korobilis (2012) such that $\kappa = 0.96$ and $\lambda = 0.99$. At last, the initialised boundary condition for all models evaluated is expressed by $\beta_0^{(k)} \sim N(0, 4I^{(k)})$, where $I^{(k)}$ denotes the identity matrix subject to the total entries number $\beta_t^{(k)}$.

STUDENT & UNIT DETAILS – TO BE COMPLETED BY THE STUDENT

Department of Economics

UNIVERSITY OF BATH

Surname, Given Names	Solomou Marina
E-mail address	Marinasolomou96@gmail.com
Phone number (optional)	
Student ID	1100670730
Candidate number	04776
Unit Name and Code	ES50078
Unit Convenor	Chris Martin
Student's Department	Department of Economics

Question	**attempted**	**Question**	**attempted**
1)		**10)**	
2)		**11)**	
3)		**12)**	
4)		**13)**	
5)		**14)**	
6)		**15)**	
7)		**16)**	
8)		**17)**	
9)		**18)**	

Unit Convenor	**Moderator**	**Final Mark**

EXAM CODE (ESXXXXX)

DECLARATION – PLEASE READ CAREFULLY

When you enrolled as a student at the University of Bath, you agreed to abide by the University's rules and regulations, and agreed that you would access and read your programme handbook. This handbook contains references to, and penalties for, unfair practices such as collusion, plagiarism, fabrication or falsification. The University's Quality Assurance Code of Practice, QA53 Examination and Assessment Offences, sets out the consequences of committing an offence and the penalties that might be applied. These penalties will continue to apply to Semester 2 assessments in the academic year 2019/20 and will not be affected by changes to the Regulations for Students during the COVID-19 disruption.

By submitting the attached assessment, you confirm that:

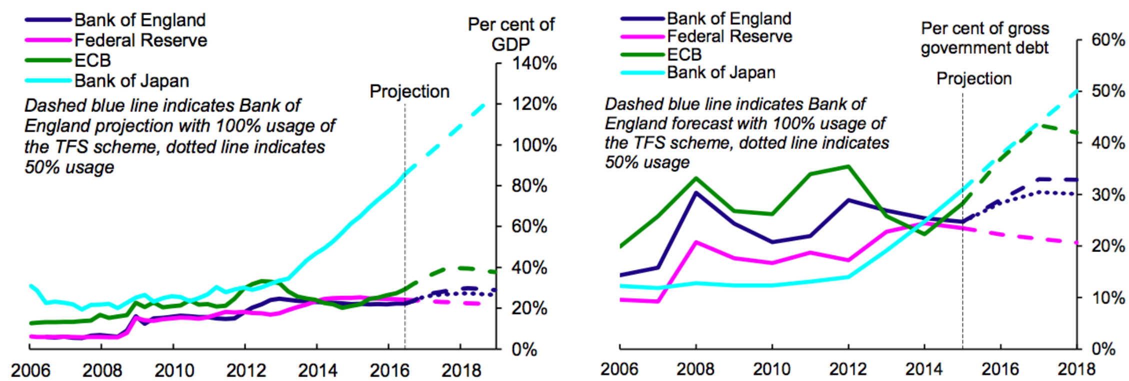
1. You have not impersonated, or allowed yourself to be impersonated by, any person for the purposes of this assessment.
2. This assessment is your original work and no part of it has been copied from any other source except where due acknowledgement is made.
3. You have not previously submitted this work for any other unit/course.
4. You give permission for your assessment response to be reproduced, communicated, compared and archived for plagiarism detection, benchmarking or educational purposes.
5. You understand that plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to disciplinary action.
6. You understand that this assessment is undertaken without invigilation, and that you have not communicated with and will not communicate with anyone concerning this assessment until the deadline for submission.
7. No part of this assessment has been produced for, or communicated to, you by any other person.

1. Explain the portfolio substitution and bank lending channels of Quantitative Easing. Discuss whether the Bank of England should extend the use of Quantitative Easing as part of its response to the Coronavirus Pandemic.

Quantitative easing (QE) is one of the primitive means policy-makers employed for the revitalisation of the economic performance during the aftermath of the Great Financial Crisis which erupted on Monday 15 September 2008, following the bankruptcy of the Lehman Brothers (Borio & Zabai (2016)). In the years that followed, trading activity in financial markets around the world became immobilised. Major economies were faced with a substantial increase in unemployment and a sharp fall in output due to contractions in the supply of credit to the private sector caused by disruption of the financial sector (Martin & Milas (2012*b*)). Hence, the immediate response of central banks of advanced economies included a sharp reduction in policy rates to historically low levels to soothe the impact on markets, offset the downfall in demand and re-stabilise the economy. Although a total economic collapse had successfully been prevented, it was not long until the effects of the turbulent market conditions were reflected by the sluggish recovery rate of the real economy (Yeyati et al. (2009)). Notwithstanding the ineffectiveness of ordinary policy instruments to impose a turnaround in the economic performance, and with central banks facing the zero lower bound constraint, policy-makers announced the exercise of unexplored policy operations targeting the stimulation of economic activity via the implementation of various expansionary monetary policies, one of which is known as Quantitative Easing.

The current essay addresses two points, first of which is the provision of an insightful entry node attributed to the burgeoning literature regarding the portfolio substitution and bank lending transmission channels of QE in real economy and secondly the evaluation of the effectiveness of the so-called "unconventional" monetary policy in successfully easing financial conditions and supporting growth during times of financial distress. Finally, based on the underlined discussion, this essay provides an informed justification on whether an additional round of QE by the Bank of England (BoE) would favour economic advancement following the impact of the Coronavirous pandemic outbreak in real economy.

The concept of QE was pioneered by the Bank of Japan (BoJ), when a member of the Policy Board suggested its implementation targeting monetary base and awakening the country's persistently low inflation (Hoshi & Kashyap (1999)). Large scale asset purchase programmes (LASP) were later on adopted by central banks worldwide as an explicit of loosening monetary policy following the global financial meltdown. The purchase programmes increased the central banks reserves, beyond the required level to hold the policy at zero rate, with magnitudes between $20 - 40\%$ relative to the overall stock of government debt, and $30 - 80\%$ of nominal GDP as illustrated by Figure 1, thus the term "quantitative". The BoE enforced the first expansion of its balance sheet in 2009, completing a LASP of £200 million mainly composed of medium-to-long-term UK government bonds, which was later on expanded to the total cumulative value of 357 million, $\approx 25\%$ of the country's GDP, focusing on the injection of liquidity in the economy by expanding money supply base and stimulating domestic demand aiming to meet the 2% CPI inflation target (Bernanke et al. (2004)). Additionally, the policy aimed the devaluation of the currency and augmentation of the entity of interest rates by lowering borrowing costs as an impulse to ease economic conditions for both firms and households supporting output (Martin & Milas (2012*a*)).



Source: Bank of England, Federal Reserve, Bank of Japan, European Central Bank, Bloomberg, Thomson Reuters DataStream and Bank calculations. *Source: Bank of England, Federal Reserve, Bank of Japan, European Central Bank, Bloomberg, Thomson Reuters DataStream and Bank calculations.*

Figure 1: Graphical Representation of QE Transmission Mechanism Channel according to Joyce et al. (2011).

The transmission mechanism of QE is composed of two branches: (i) the expansion of the central bank's balance sheet, creating bank reserves and increasing broad money holdings targeting the purchase of long-term bills; (ii) the swap of the underlined bills for longer-term bonds known as maturity extension programme (Haldane et al. (2016)). However, according to macroeconomic models neither of the branches should affect economic performance since expanding the reserve supply and keeping a relative low opportunity cost hardly affects the economy. Additionally, asset re-allocations between public and private sector portfolios has no substantial economic impact if no constraints are imposed on the positions of investors' portfolios other than budget constraints (Woodford (2012)).

Therefore, to fully grasp the idea of QE propagation in the real economy, two categories of market frictions are imposed namely information frictions and market frictions, employing a more realistic representation of market conditions. The former arises from the information asymmetry between market participants either regarding the future course of the macroeconomy or future policy response actions of the monetary authorities and the latter arises due the varying degrees of substitutability of different asset classes and specific preferences of market investors regarding assets with particular characteristics giving rise to market imperfections. This friction group determines the portfolio balance, bank lending channels. Hence, QE is classified as a state-dependent policy, whereas its effectiveness jointly depends on various parameters including the design of the LASP, the assets purchased, the structure of the financial system, different friction sets and the portfolio allocation decisions of financial intermediaries. According to evidence, QE is increasingly operative the weaker the economic and financial state of a system is (Haldane et al. (2016)).

Presumably, the most advocated transmission channel amongst central banks, is the portfolio substitution channel initially introduced as inter alia (Aschheim (1973)). The underlying theory suggests a direct effect on different asset yields by central bank interventions as a consequence of imperfect asset substitutability and investors heterogeneity. Put differently, arguing against the Ricardian Equivalence theory (Bernheim (1987)), by relaxing the assumption of investors' rationality and implementing credit restrictions to account for a tightening in market liquidity, reducing the asset's market availability reduces the stock of long-dated assets privately held. Consequently, the decreased duration risk to hold in the aggregate lead market participants to demand a lower premium resulting in the reduction of long-dated assets premia inflating the asset's price and depressing its yield (see Appendix 1). Correspondingly, given that cash is an imperfect substitute to assets, investors with unchanged liabilities are forced to re-balance their portfolios by purchasing assets with similar characteristics such as duration and credit risk due to the fact that short-dated bank deposits generate no profits. Hence, this mechanism bids up prices of similar asset classes, lowers borrowing costs increasing the net wealth of investors by easing credit conditions and enables companies to raise credit funds easier. As a result, households' capital gains increase promoting nominal spending and stimulating aggregate demand as shown by Figure 2.

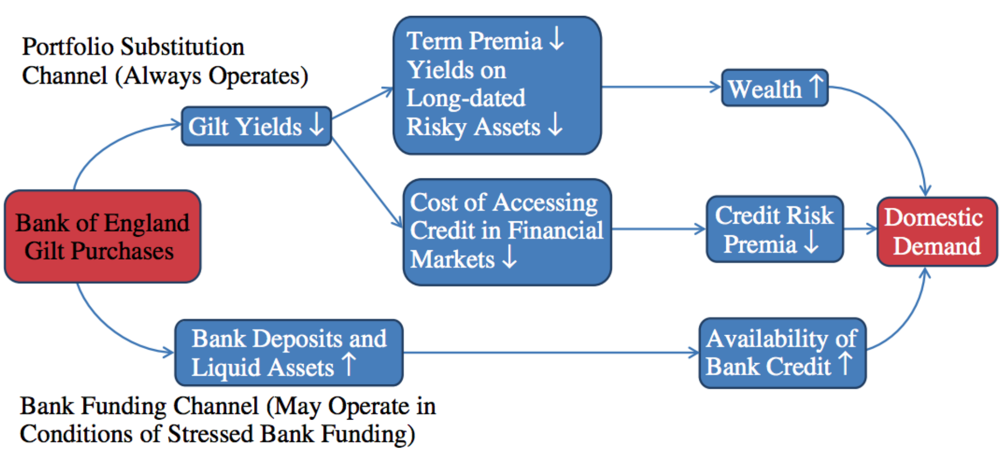


Figure 2: Graphical representation of BoE's Gild purchases impact on Domestic Demand (Joyce et al. (2012)).

However, accounting for all different types of market participants, such as Pension Funds (PF) with longer maturity investments and entrepreneurs with short-term equities, in the case where no perfect alternatives are available in the market, some investors might choose to abstain from re-investing the capital received from selling their assets to the central bank, and instead choose to hold cash.

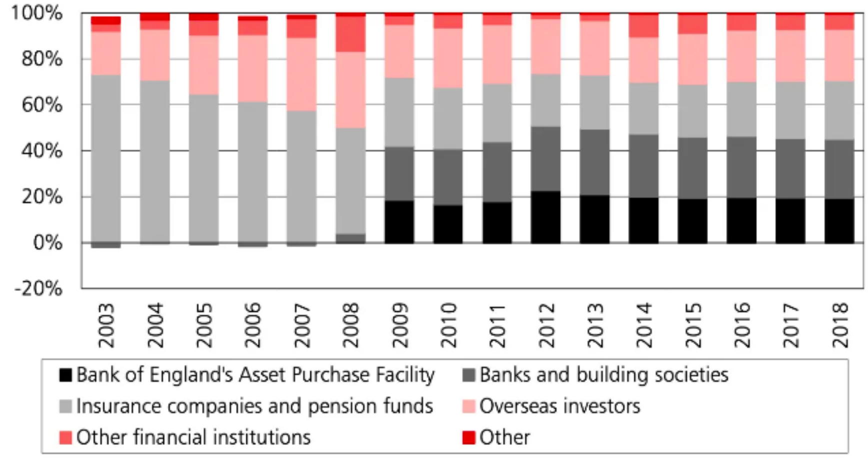


Figure 3: Chart representation of increase in gilt holdings of BoE during the period of 2009-2010, (Source: HM Treasury UK debt report 2019/20).

This might result in the further tightening of market liquidity giving rise to a liquidity trap. Specifically, the BoE would complete a LASP of longer-dated government bonds from the domestic non-financial sector including insurance companies and pension funds (ICPFS), as illustrated by Figure 3, indirectly through a commercial bank acting as a finance intermediator, funded by the creation of the commercial bank's reserves - essentially issuing a zero duration risk, zero credit risk instrument (Haldane et al. (2016)). Consequently, the deposits of the ICPFS held at the commercial bank enlarge whilst bond holdings decrease, shifting the ICPFS' asset composition from longer to shorter dated asset holdings as illustrated by Table 1. This mechanism reduces their term premium encouraging the re-investment in shares of sovereign securities, equities and corporate bonds as illustrated by Figure 4 (Shah & Malki (2018)).

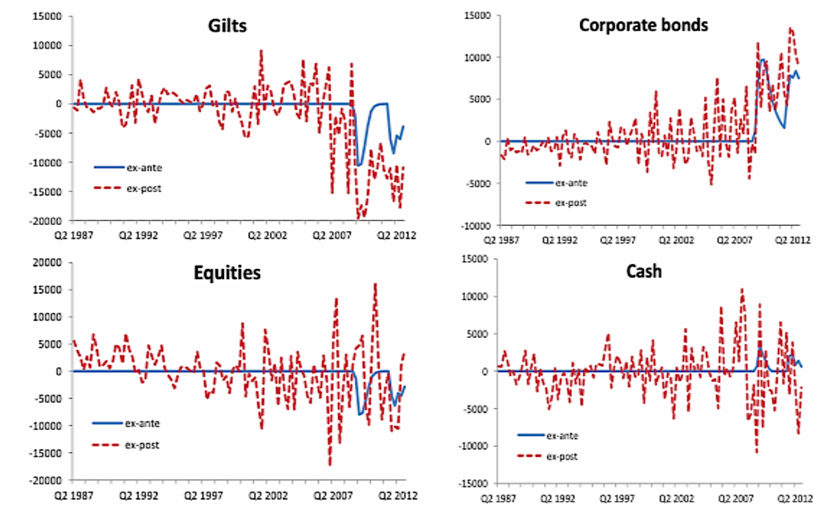


Figure 4: Graphical representation of QE's impact on UK's ICPFS, ex-ante and ex-post effects given in, £millions (Joyce et al. (2011)).

Assets	Liabilities
Deposits at commercial banks ↑	Pension liabilities
Bonds ↓	

Table 1: Effect of Portfolio Substitution Channel on ICPFS Balance Sheets.

The second transmission channel investigated includes the bank lending channel and for its appropriate evaluation the introduction of the money multiplier m, is presented as an intuitive first step (see Appendix 2).

$$m = \frac{Broad\ Money}{Narrow\ Money} = \frac{M1}{M0} = \frac{1}{cc + rr(1 - cc)} \ , \tag{1}$$

where cc signifies the fraction of cash held by the private sector proportional to the broad money supply M_0, and rr denotes the actual ratio of commercial bank reserves to deposits, held by the Central Bank. As already mentioned, in the case where the BoE finances LASPs from the non-bank financial sector through the creation of reserves, the commercial bank's balance-sheet composition changes, and liabilities are matched with new reserves representing the on-balance-sheet assets, illustrated by Tables 2 (McLeay et al. (2014)). Thus the the ratio of loans to reserves decreases resulting in the reduction of commercial banks profitability (see Appendix 3).

Assets	Liabilities
Reserves at BoE ↑	Deposits ↑
Loans	Equity
Other Assets	

Table 2: Effect of Bank Lending Channel on Commercial Banking System.

Hence, returns-driven banks are faced with a trade-off between profitability and credit risk. In the case where banks are less vulnerable to insolvency they are incentivised to increase lending aiming higher profits and re-balance their balance-sheet compositions to pre-LASPs levels. Consequently, by issuing new loans, commercial banks create new broad money on the household's balance-sheets in terms of bank deposits, referred to by economists as "fountain pen money". Alternatively, during times of financial distress, the banking sector may prioritise holding liquid assets such as reserves over profits. Recalling equation (1), under the assumption of constant money holdings by firms and households, with increasing bank reserves holding m unchanged, a multiplied increase in broad money is suggested giving rise to potential hyperinflation. However, the underlying scenario did not occur as commercial banks chose to hold an excessive amount of reserves resulting to a substantial drop of UK's money multiplier as illustrated by Figure 5.

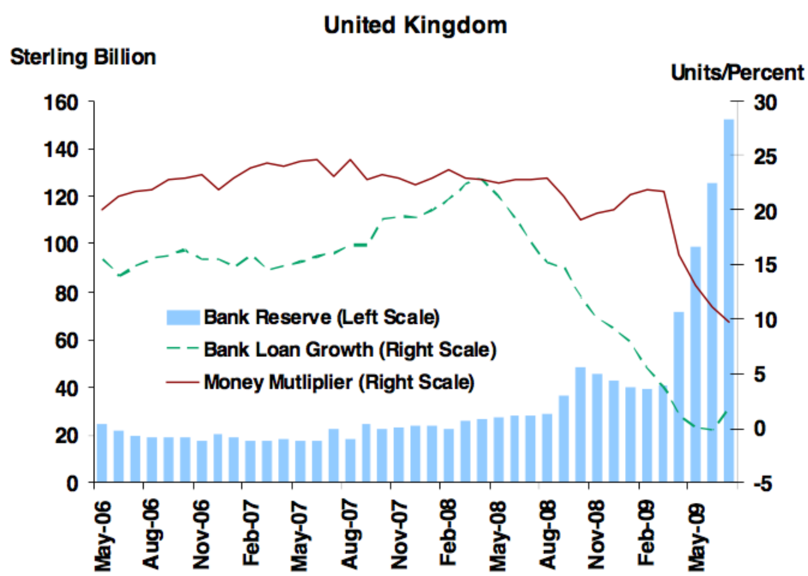


Figure 5: Graphical representation of UK's Bank Reserves, Bank Loan Growth and Money Multiplier over the period of May 2006 until May 2009 (Disyatat (2011)).

The efficacy determination of the investigated monetary policy is a rather trivial task as it is exercised under severe macroeconomic conditions. Therefore, the main focal point of this section includes the analysis of empirical evidence in the framework of former QE programmes of the UK, examining differed macroeconomic factors that quantify its significance. Under this scope, various counterfactual studies, event studies and more advanced time series regressions are employed and discussed. Joyce et al. (2010), quantifies the impact of QE on gilt yields and Overnight Index Swap rates (OIS) by assessing their intermediate reaction to six pieces of QE announcements, as the majority of impact occurs due to the formation of the market participant's expectations. Specifically, the study focuses on examining the direct reaction of the market via the observation of market prices movements over a 2-day window, following the release of related news. The reported estimates suggest that the announcement made on March 2009 of £75 *billion* gilt purchases with residual maturities ranging between $5-25$ years resulted in the total yield reduction of these gilts of 80 basis point (bps). Accordingly, their OIS rates did not exhibit a similar fall indicating that the causation of this event was not due to the expectations of short-term rates decrease but rather on the portfolio substitution channel. Additionally, Meier (2009) quantifies the same effect estimating an impact of approximately $40-100$ bps stating that the"effect of QE on guilt yields of, at any very least, 35-60 basis points". Oppositely, the finding of Meaning & Zhu (2011) and Glick & Leduc (2012) regarding QE1 suggest a smaller impact of roughly 50 bps. Accordingly, regarding the 2 following QE programmes, Meaning & Zhu (2011) state that no impact was recorded on the yield curve of government bond rates, but instead a small rise was identified. Finally, according to the 1-day window event study of Martin & Milas (2012a), findings suggest that QE3 resulted in the 7 bps reduction in the 20-year rate, 5 bps increase in the 10-year rate and 10 bps drop in the 5-year rate. To summarise, the underlying studies signify the efficacy of QE in decreasing UK government bond yields.

The second type of QE econometric study includes the investigation of bond rates using historical data from pre-crisis and post-crisis, quantifying the impact of variations in government bond holdings. Specifically, Joyce et al. (2011), examines the portfolio-rebalance channel by employing a VAR model and introducing system shocks representing the fall in government bonds availability to the private sector. The associated findings suggest QE1 compressed the yield curve by 85 bps. Similar findings with varying sizes where deduced by Glick & Leduc (2012) and Meaning & Zhu (2011). Finally, more sophisticated models developed by Hamilton & Wu (2012) estimate the average decrease of the bond yield to be 13 bps.

The impact of QE as defined by Martin & Milas (2012a), is a causation chain connecting government bond rates, assets returns, aggregate demand and accordingly inflation and output. This part focuses on the latter part of this chain. The sizable impact of QE on corporate bond rates is investigated by the event study of Joyce et al. (2011) with findings indicating QE1 announcements resulted in the decline of the UK BBB corporate bonds yield curve during the time window of $4^{th}-6^{th}$ of March 2009 signifying the early purchases of corporate bonds by investors awaiting purchases of their government bond. Additionally, he investigates the equity yields response reporting a fall of 282 bps. Glick & Leduc (2012) address the effect of QE1 reporting the currency depreciation of magnitude $1.5-3.5\%$.

According to the BoE, the lending channel was expected to impose a continuously growing effect in the wider economy. However, Breedon et al. (2012) reports that over the QE period, the total lending activity excluding loan transfers and securitisations experienced a fall of £197.5*billion*, stating that the underlying phenomenon is attributed to equity issuance by the UK and the significantly increased level of new debt. Conducting a separate analysis, Butt et al. (2014) employs an instrumental variable for the investigation of the impact of increasing deposits on lending, where he reports that no statistically significant evidence was found supporting the aforementioned argument. Nonetheless, the underlying study reports that the portfolio re-balancing channel gives rise to "flight deposits" devaluing the bank lending channel concluding that the implementation of QE resulted in no significant increase in lending. Hence, one can observe the undoubtedly weak performance of the bank lending channel which lead to the Funding for Lending Scheme implementation in 2012. Finally, the impact of QE on inflation and output was examined via long term interest rates, by numerous econometric studies including the one of Baumeister & Benati (2010), who observes a powerful impact in both inflation and output by the spread compression, predicting that in the absence of QE, inflation would drop below -4% and output growth would have experience a drop to -14% instead. Kapetanios et al. (2012), employing a similar approach obtains more moderate estimates, concluding that QE1 gave rise to inflation and output by $75 - 150$ bps and $150 - 200$ bps respectively.

Based on the presented event studies initial LASP successfully decreased UK's government bond rates particularly on the longer end of the yield curve. Additionally, econometric estimates of QE's impact on inflation and output proves its efficiency in minimising the downturn magnitude resulted by the Great Financial Crisis by preventing further declines in inflation and output than the ones expected. However, evidence questions the perseverance of these effects in the long-run signified by the limited evidence of succeeding QE programmes imposing adequate effect in easing financial conditions, suggesting that QE solely is not powerful enough to exhilarate economic recovery. In light of this, with the UK's economy facing the catastrophic impact of the Coronavirus Pandemic on the real economy, an additional QE round should be supported as part of BoE's response. However, in order to target long-term economic effects alternative policy instruments should be employed, such as a negative Interest rates policy or Helicopter Money.

$word\ count : 2,470$

2. Explain the concept of an inflation bias in monetary policy. How effective is Central Bank independence as a solution to the inflation bias?

According to literature, during the twentieth century most countries' economies have been characterised by dynamic inconsistencies and positive inflation rates, arising from the counter-cyclical interactions between an activist monetary authority and rational private agents, governed in the context of the expectation-Augmented-Phillips curve. (Cukierman & Gerlach (2003)). Specifically, under this discretionary regime, the monetary authority has the freedom to revise its monetary instrument, generating more inflation than the private agent's expectations by exercising expansionary monetary policies such as printing more money. Consequently, the surprise inflation may result in the revitalization of economic activity, reducing at the same time the government's nominal liabilities. However, due to the non-cooperative nature of the game between the two parties, these elements of surprises and the resulting benefits cannot systematically arise in equilibrium. After a certain point, economic agents realise the incentives of policy-makers and obliterate patterns of consistent surprises by revising their inflation expectations. As a result, inflation shocks are generated in the economy, reflected in equilibrium as excessive rates of both money growth and average inflation than otherwise giving rise to inflation bias (Barro & Gordon (1983)). The main focal points of the current essay therefore are the outline of the circumstances out of which the inflation bias arises including the formalisation of the economic mechanism behind the concept, and the evaluation of Central Bank Independence (CBI) as an alternative monetary regime for the solution of the underlined economic phenomenon.

This section formulates the main objectives of the monetary authority by adopting a framework that reduces the analysis to a pair of independent single-agent decision problems, originally developed by Barro & Gordon (1983). Initially, the features of the economic equilibrium state are identified considering the event sequence of the model:

(a) The private agents decision rule, which formulates the private sector's actions expressed as a function of all available information.
(b) The expectation function outlining expectations of the private sector also given as a function of all available information to the private sector and the according nominal wages.
(c) The policy-maker's response to any potential realised supply shock by revising the behaviour of policy instruments, denoted as a function of the available information set to the monetary authority.

The game's outcome is only considered to be a rational expectations equilibrium if the decision rule of the private sector initialised in (a) is optimum, presuming that its expectations are estimated at (b), and it is also optimal for policy-makers whose response is determined at (c) in accordance with the private sector's expectations (b), under the assumption that the monetary authority observes the private sector's decision rule (a). Under this counter-cyclical regime, it is assumed that private agents commit to the determination of the nominal wage prior to the pre-commitment of the policy-maker to the growth rate of the nominal money supply. Consequently, by acting in a different manner to the private sector's expectation the policy-maker is faced with the opportunity of inflation surprise after the private sector agrees nominal contracts. Therefore, encountered with a maximising monetary authority, the private sector cannot maintain any form of expectations knowing that the policy-maker is incentivised to deviate from it. As a result, the realisation of expectations equilibrium is not subject to any prior constraints.

More specifically, the policy-maker's optimisation problem is expressed in this case as the the maximisation of a utility function U, depending on the parameters of both economic inflation π, and output y as illustrated by Equation (2). Under the discretion regime, this is achieved by revising the stabilisation policy instrument Δm, denoting the growth rate of the money supply every time period.

$$\max_{\Delta m} U = \big\{\lambda(y - y_n) - \frac{1}{2}\pi^2\big\}, \tag{2}$$

where y_n represents the economy's natural output rate, and λ signifies the relative weight attributed by the monetary authority on output expansion relative to the stabilisation of inflation. Observing the optimisation equation form, output enters linearly in the relation. For a constant marginal utility, more output is preferred to less, and inflation appears to enter quadratically inducing an increasingly marginal dis-utility. Hence, the objectives of the policy-maker under this regime are derived which appear to be targeting zero inflation rate and maximum economic output.

However, for the identification of the optimal policy response, the optimisation problem is expressed in terms of the policy instrument Δm. The representation of the economy is employed from the analysis of Barro & Gordon (1983). Firstly, the supply side of the economy is denoted by an expectations-augmented Phillips curve, reflecting a proxy of the overall state of real economic activity,

$$y = y_n + \alpha(\pi - \pi^e) + \epsilon \quad , \tag{3}$$

where π^e denotes the inflation rate expectations of the private sector, α is a positive parameter demonstrating the effect of unanticipated inflation on employment, also known as the Phillips curve gradient, and ϵ is an independent and identically distributed (i.i.d) random element with zero mean, denoting stochastic aggregate supply shocks. Where the inflation gap $(\pi - \pi^e)$ is positive, firms expand employment depressing real wages. Alternatively, where the expected inflation exceeds the actual inflation rate, driving the inflation gap to be negative, employment will decline with real wages exceeding the level of expectations. Hence, as inflation costs arise solely from inflation expectations whilst surprise inflation induces economic activity, the monetary authority only benefits from the generation of unexpected inflation. Similarly, the supply side of the economy is expressed by the Aggregate Demand (AD) equation:

$$\pi = \Delta m + \nu \quad , \tag{4}$$

where ν denotes possible aggregate demand shocks in the economy also i.i.d with zero mean known as the velocity shock. Substituting Equation (4) in Equation (3) the Aggregate Demand (AS) curve is expressed as follows

$$y - y_n = \alpha(\pi - \pi^e) + \epsilon = \alpha(\delta m + \nu - \pi^e) + \epsilon. \tag{5}$$

Finally, substituting the obtained expression in Equation (3), the optimisation objective is expressed as a function of Δm.

$$U = \lambda(\alpha(\Delta m + \nu - \pi^e) + \epsilon) - \frac{1}{2}(\Delta m + \nu)^2 \tag{6}$$

Taking the first order partial differential with respect to Δm, the optimality condition and the solution to the policy authority's response objective is:

$$\frac{\partial U}{\partial \Delta m} = 0 \Longrightarrow \alpha\lambda - \Delta m - \nu = 0 \Longrightarrow \Delta m + \nu = \alpha\lambda. \tag{7}$$

Recalling the event sequence of the game, policy-makers do not realise the policy shock ν, therefore taking its expected value $E\nu = 0$ and substituting it into Equation (7), the optimal policy response appears as the product of the aggregate supply relation slope α with the sensitivity parameter of U on the output gap, λ.

$$\Delta m = \alpha\lambda \tag{8}$$

Substituting the realised solution into the Equation (4), inflation is expressed as a function of the policy instrument Δm

$$\pi = \Delta m + \nu \longrightarrow \pi = \alpha\lambda + \nu \tag{9}$$

Finally, taking expectations for the estimation of inflation expectations,

$$\pi^e = \alpha\lambda + E(\nu) = \alpha\lambda \quad , \tag{10}$$

the average rate of inflation characterising the private sector's expectations equals $\alpha\lambda$. The prediction power of the underline expression is limited as it fails to forecast possible deviations of inflation from the expected value that may arise due to potential policy shocks. Hence, π can randomly but no systematically diverge from π^e, the same way that monetary policy cannot systematically mislead the private sector.

In the perspective of economic output given by,

$$y = y_n + \alpha\underbrace{(\pi - \pi^e)}_{\nu} + \epsilon = y_n + \alpha\nu + \epsilon \ , \tag{11}$$

One can observe that deviations from equilibrium solely depend on policy shocks and supply shocks signifying the inability of policy-makers to affect economic output since average inflation if fully anticipated by market participants,$\pi_e = E[\Delta m] = \alpha\lambda$, illustrated by Figure 1. Hence, under the baseline model of discretionary policy regime, inflation is expressed as:

$$\pi \ = \ \underbrace{\alpha\lambda}_{inflation\ bias} \ + \ \epsilon \tag{12}$$

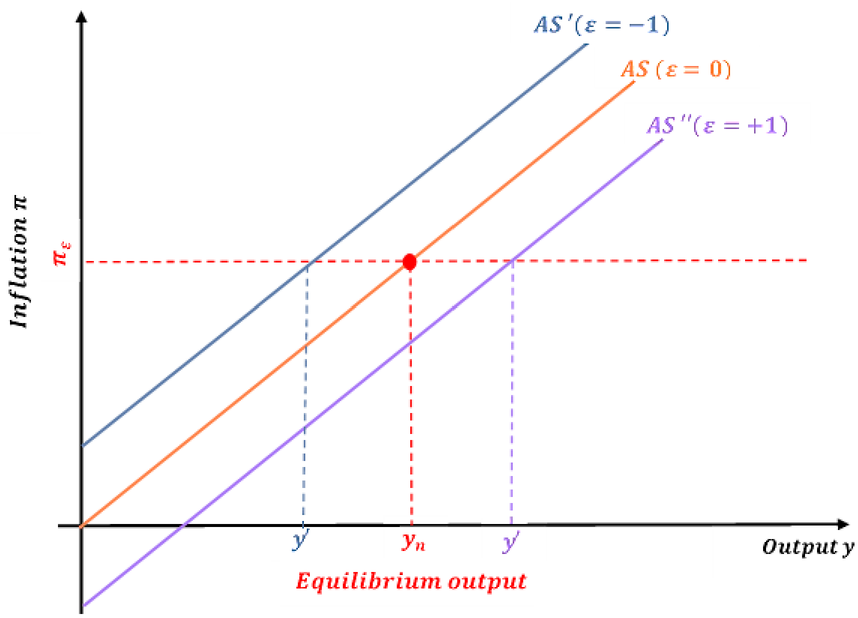


Figure 1: Graphical Representation of Aggregate Supply Relationship (AS) shifts according to different realised supply shocks $\epsilon = -1, 0, +1$ and their corresponding equilibrium output for the case whre $\pi = \pi^{\epsilon}$.

Equilibrium suffers from positive average inflation bias $\alpha\lambda$, which increases proportionally to the attributed weight by the central bank on the output objective λ as well as the marginal benefit of excess output governed by the money surprise element α, providing no desirable effects of higher output. The underlying phenomenon can be explained by two offsetting incentives of policy-makers which give rise to the output-inflation trade-off. Specifically, according to Equation (11), by imposing a greater money growth rate, for a specific π^e value, gives rise to the inflation gap and consequently achieve a greater economic output. However, such practice increases inflation, suppressing the optimisation function. Re-arranging Equation (11) for output gap and substituting it into the optimisation problem, becomes:

$$\underbrace{y - y_n}_{output\ gap} = \alpha \underbrace{(\pi - \pi^e)}_{inflation\ gap} + \epsilon \tag{13}$$

$$U = \lambda \underbrace{(y - y_n)}_{output\ gap} - \frac{1}{2}\pi^2 = \lambda(\alpha \underbrace{(\pi - \pi^e)}_{inflation\ gap} + \epsilon) - \frac{1}{2}\pi^2 \tag{14}$$

Imposing the optimality condition by taking the partial difference w.r.t inflation, the optimal inflation rate for the monetary authority is:

$$\frac{\partial U}{\partial \pi} = \underbrace{\alpha\lambda}_{marginal\ benefit} - \underbrace{\pi}_{marginal\ cost} = 0 \tag{15}$$

$$\longrightarrow \pi^* = \alpha\lambda \ , \tag{16}$$

where the marginal benefit signifies the benefit to the policy-maker of inflation in terms of greater output as illustrated in Figure 2.

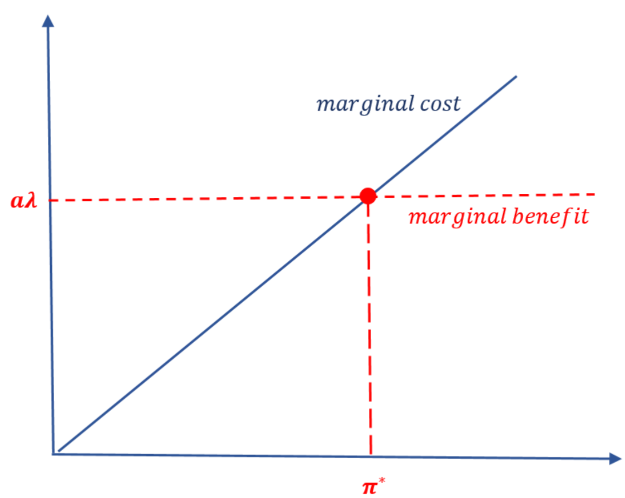


Figure 2: Graphical Representation equilibrium inflation π^* employing a linear objective function, under the discretion regime.

Therefore, by construction of the problem, the necessity of revising the policy-makers incentives is emphasised, motivating the employment of a strategy where the policy-maker commits to an action enforcing a specified policy rule such that the only rational expectations for the private sector to hold are the ones shaped by the underlying rule giving rise to the debate between "discretion-versus-commitment", and consequently to the possibility of a superior economic outcome. The exploitation of such rule-enforcement-regime is underlined by the following alternative model developed upon the key assumption of output stabilisation around a specified level. Under this model, the objective function of the policy-maker is expressed as the minimisation of a loss function L, in contrast to the maximisation of a utility function U, where L is expressed as:

$$L = \frac{\lambda}{2}(\underbrace{\underbrace{y - y_n}_{\textit{output gap}}{}^{2} - \kappa)}_{\textit{stabilisation term}} + \frac{1}{2}\pi^2 \tag{17}$$

observing the current optimisation problem, the difference between output gap and κ, and inflation enter quadratically in the equation signifying increasing marginal dis-utility of the monetary authority in the cases where output deviates from the stabilisation term $y - y_n - \kappa$ and inflation differs from 0. Equivalently, policy-makers objectives are reduced to zero inflation and output stabilisation around $y_n + \kappa$. Once again, expressing the the objective function in terms of policy instrument Δm, the following equation is deduced:

$$L = \frac{\lambda}{2}\left(\alpha(\Delta m + \nu - \pi^e) + \epsilon - \kappa\right)^2 + \frac{1}{2}(\Delta m + \nu)^2 \tag{18}$$

Imposing the optimisation condition and partially differentiating the objective w.r.t the policy instrument Δm, the policy instrument is required to satisfy:

$$\frac{\partial L}{\partial \Delta m} = 0 \tag{19}$$

$$\longrightarrow \frac{\partial L}{\partial \Delta m} = \alpha\lambda\left(\alpha(\Delta m + \nu - \pi^e) + \epsilon - \kappa\right) + \Delta m + \nu = 0 \;\; || \; Ev = 0, \tag{20}$$

$$\longrightarrow \alpha\lambda\left(\alpha(\Delta m - \pi^e) + \epsilon - \kappa\right) + \Delta m = 0 \tag{21}$$

Expanding Equation (21), and re-arranging for Δm:

$$\alpha^2\lambda\Delta m - \alpha^2\lambda\pi^e + \alpha\lambda(\epsilon - \kappa) + \Delta m = 0 \tag{22}$$

$$\longrightarrow (1 + \alpha^2\lambda)\Delta m = \alpha^2\lambda\pi^e + \alpha\lambda(\kappa - \epsilon) \tag{23}$$

$$\longrightarrow \Delta m = \frac{\alpha^2\lambda\pi^e + \alpha\lambda(\kappa - \epsilon)}{(1 + \alpha^2\lambda)} \tag{24}$$

One can see that the optimal money growth depends on both expected inflation and supply shock, providing the monetary-authority with a different optimal choice. Thus, accounting for the fact that policy-makers observe the π^e value set by the private sector, prior to determining Δm, the expected inflation is:

$$\pi = \Delta m + \nu = \frac{\alpha^2\lambda\pi^e + \alpha\lambda(\kappa - \epsilon)}{(1 + \alpha^2\lambda)} + \nu \tag{25}$$

Taking expectations, imposing the constraint $E[\nu] = [\epsilon] = 0$, expected inflation becomes:

$$\pi^e = \frac{\alpha^2\lambda\pi^e + \alpha\lambda\kappa}{(1 + \alpha^2\lambda)} \tag{26}$$

$$\Longrightarrow (1+\alpha^2\lambda)\pi^e = \alpha^2\lambda\pi^e + \alpha\lambda\kappa \tag{27}$$

$$\longrightarrow \pi_e^* = \underbrace{\alpha\lambda\kappa}_{\textit{inflation bias}} \tag{28}$$

Substituting Equation (27) into the optimal monetary policy, Equation (24):

$$\Delta m \;=\; \alpha\lambda\kappa \;-\; \underbrace{\frac{\alpha\lambda}{1+\alpha^2\lambda}\epsilon}_{\textit{response to supply shocks}} \tag{29}$$

Taking expectations, the expected rate of money growth equals:

$$\Delta m = \alpha\lambda\kappa \propto \kappa \quad , \tag{30}$$

such that the inflation gap becomes:

$$\pi - \pi^e = \underbrace{\Delta m + \nu}_{\pi} - \underbrace{\alpha\kappa}_{\pi^e} \tag{31}$$

$$\longrightarrow \alpha\lambda\kappa - \frac{\alpha\lambda}{1+\alpha^2\lambda}\epsilon + \nu - \alpha\lambda\kappa = -\frac{\alpha\lambda}{1+\alpha^2\lambda}\epsilon + \nu \tag{32}$$

Substituting equation into Equation (3), output is expressed as:

$$y = y_n + \alpha(\pi - \pi^e) + \epsilon = y_n + \alpha\left(-\frac{\alpha\lambda}{1+\alpha^2\lambda}\epsilon + \nu\right) + \epsilon \tag{33}$$

Re-arranging for y:

$$y = y_n + \alpha\nu + \left(\frac{1+\alpha^2\lambda - \alpha^2\lambda}{1+\alpha^2\lambda}\right)\epsilon = y_n + \alpha\nu + \frac{1}{1+\alpha^2\lambda}\epsilon \tag{34}$$

The monetary authority cannot affect the output movement as once again as it depends on both ν and ϵ, therefore no impact in inflation can be imposed, giving rise to inflation bias. For the graphical illustration of the solution, Equation (25) is partially differentiated w.r.t. π^e:

$$\frac{\partial\pi}{\partial\pi^e} = \frac{\alpha^2\lambda}{1+\alpha^2\lambda} < 1 \tag{35}$$

denoting, that in equilibrium, point A, inflation is positive regardless of the value of expected inflation as illustrated by Figure 3.

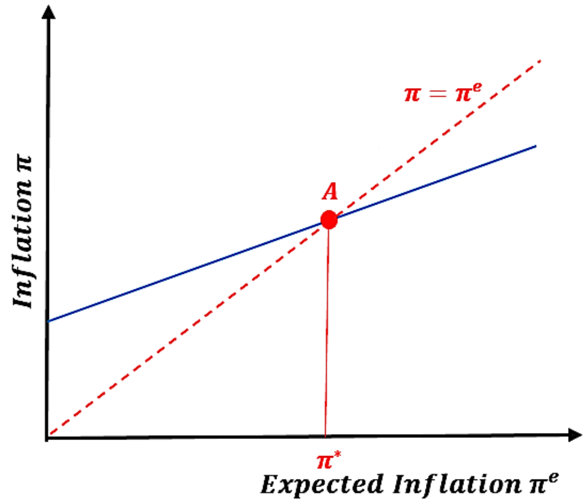


Figure 3: Graphical Representation inflation as a function of expected inflation π^* employing a linear objective function, under the discretion regime.

When however, policy-makers commit to an action enforcing a policy rule keeping Δm constant, then the only rational expectations for the private sector to hold are the ones shaped by the underlying rule. Under this regime:

$$\Delta m = 0 \longrightarrow \pi = \Delta m + \nu = \nu \tag{36}$$

Hence, expected inflation and average output become:

$$\pi^e = E[\pi] = E[\nu] = 0 \longrightarrow y = y_n + \alpha(\pi - \pi^e) + \epsilon = y_n + \alpha\nu + \epsilon \tag{37}$$

$$\Longrightarrow E[y] = y_n \quad , \tag{38}$$

achieving the same output as the optimal discretionary regime and eliminating the inflation bias. However, recalling the definition of rational expectations equilibrium and the reputational sequence of the game, $\{\pi^e = 0, \pi = 0\}$ is not considered as an equilibrium state of the model as the policy-makers are faced with the temptation to "cheat" every period to assure any potential benefit of inflation shocks arising from economic distortions. Nevertheless, such action would resonate the viability of the attained equilibrium threatening to relocate the economy back to the substandard discretion equilibrium. Accounting for the repeating nature of interactions between the two players, Barro & Gordon (1983), believe that by imposing reputational forces in the game formalisation could sustain the enforced rules, encouraging policy-makers to relinquish any possible short term gains from inflation shocks in order to secure the benefit of minimum low long-term average inflation.

Employing a multi-period analysis, where policy-makers deviate from the rule-enforcement regime such that $\{\pi^e = 0, \pi > 0\}$, the realised inflation expectations of the private-sector will be $\{\pi^e > 0\}$

in the subsequent period, giving rise to an inferior economic outcome. The underlying discussion is illustrated by Figure 4, where:

- point A signifies the baseline case under the discretion regime under the assumptions $\epsilon = \nu = 0$:

$$\pi = \Delta m + \nu = \alpha\lambda + \underbrace{\nu}_{0} \longrightarrow \pi^e = \pi = \alpha\lambda \longrightarrow y = y_n.$$

- Point B denotes the case where policy-maker commits and delivers $\Delta m = 0$,

$$\longrightarrow \pi = \pi^e = 0 \longrightarrow y = y_n. \tag{39}$$

- Point C defines the not credible case where the policy-maker deviates from the rule-enforcement regime misleading the private sector to expect $\pi = 0$, resulting in $y > y_n$.

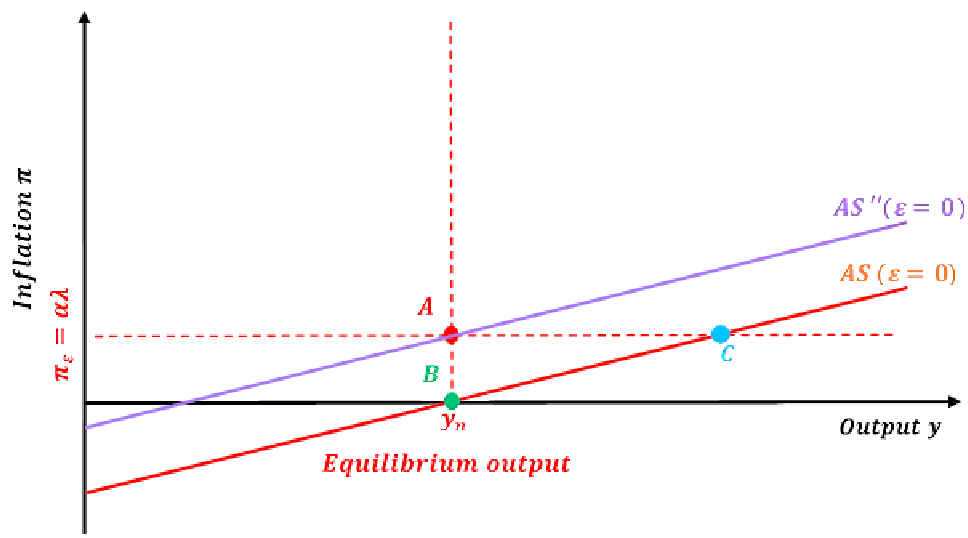


Figure 4: Graphical Representation of different economic states under the examined policy regimes, represented by the Aggregate Supply (AS) relation.

This section examines the effectiveness of CBI as a solution to the inflation bias. The optimal solution to the underlying phenomenon should be identified amongst others. Initially, the definition of a Conservative Central Banker (CCB) is examined, developed on the assumption that the monetary-authority is more inflation-averse than economic agents. Specifically, assuming that the loss function of the private sector is expressed as:

$$U = \frac{\lambda}{2}(y - y_n - \kappa^2) + \frac{1}{2}\pi^2 \ , \tag{40}$$

and accordingly the loss function of the policy-maker is denoted by:

$$U = \frac{\lambda}{2}(y - y_n - \kappa^2) + \frac{(1+\delta)}{2}\pi^2 \quad \delta > 0 \ , \tag{41}$$

one can observe the weight attributed in inflation by the policy-maker is greater than the private sector. Rogoff (1985), argues for a CCB setting monetary-policy stating that it could potentially increase social welfare. Under this regime the inflation is expressed as:

$$\pi = \frac{\alpha\lambda\kappa}{1+\delta} - \frac{\alpha\lambda\epsilon}{1+\delta+\alpha^2} + \nu. \tag{42}$$

Therefore, in the case where δ is large, inflation bias reduces substantially. However, this regime fails to provide an optimal response when the economy experiences negative supply shocks due to extreme economic conditions such as the Coronavirus Pandemic, leading the discussion to a simpler and more effective solution for the elimination of inflation bias. The output stabilisation around the equilibrium value such that $\kappa = 0$.

$$\textit{Inflation:}\ \ \pi = -\left(\frac{\alpha}{1+\alpha^2\lambda}\right)\epsilon + \nu \tag{43}$$

$$\implies E[\pi] = \pi^e = 0. \tag{44}$$

$$\textit{Output:}\ \ y = y_n + \alpha\nu + \left(\frac{1}{1+\alpha^2\lambda}\right)\epsilon \tag{45}$$

Inflation bias can be completely eliminated achieving a superior economic outcome by the policy-maker realising the structure of the economy and targeting the output stabilisation around the natural output value, since any attempt of output maximisation beyond this point gives rise to the inflation bias. Therefore, a solution for this problem includes the CBI, as commonly output maximisation is enforced by elected governments and political parties under the scope of engineering pre-election booms stimulating output growth and unemployment. The notion of CBI covers a variety of aspects such as the voting power of the government on the board, the government's role in diminishing and recruiting Central Bank's (CB) associates and the lending extend of the CB to the government. However, The current discussion emphasises two key components including the instrument and goal independence. The former outlines the degree of freedom of the CB in adjusting the policy tools according to the monetary policy's objectives and the latter refers to the ability of the CB in defining policy goals free from government interference.

Alesina & Summers (1993), investigate the relation between CBI and the variability levels of different economic variables including unemployment real interest rates and growth, employing data from $1955 - 1988$ and the sample size of 16 countries, the results of who are summarised in Figure 5.

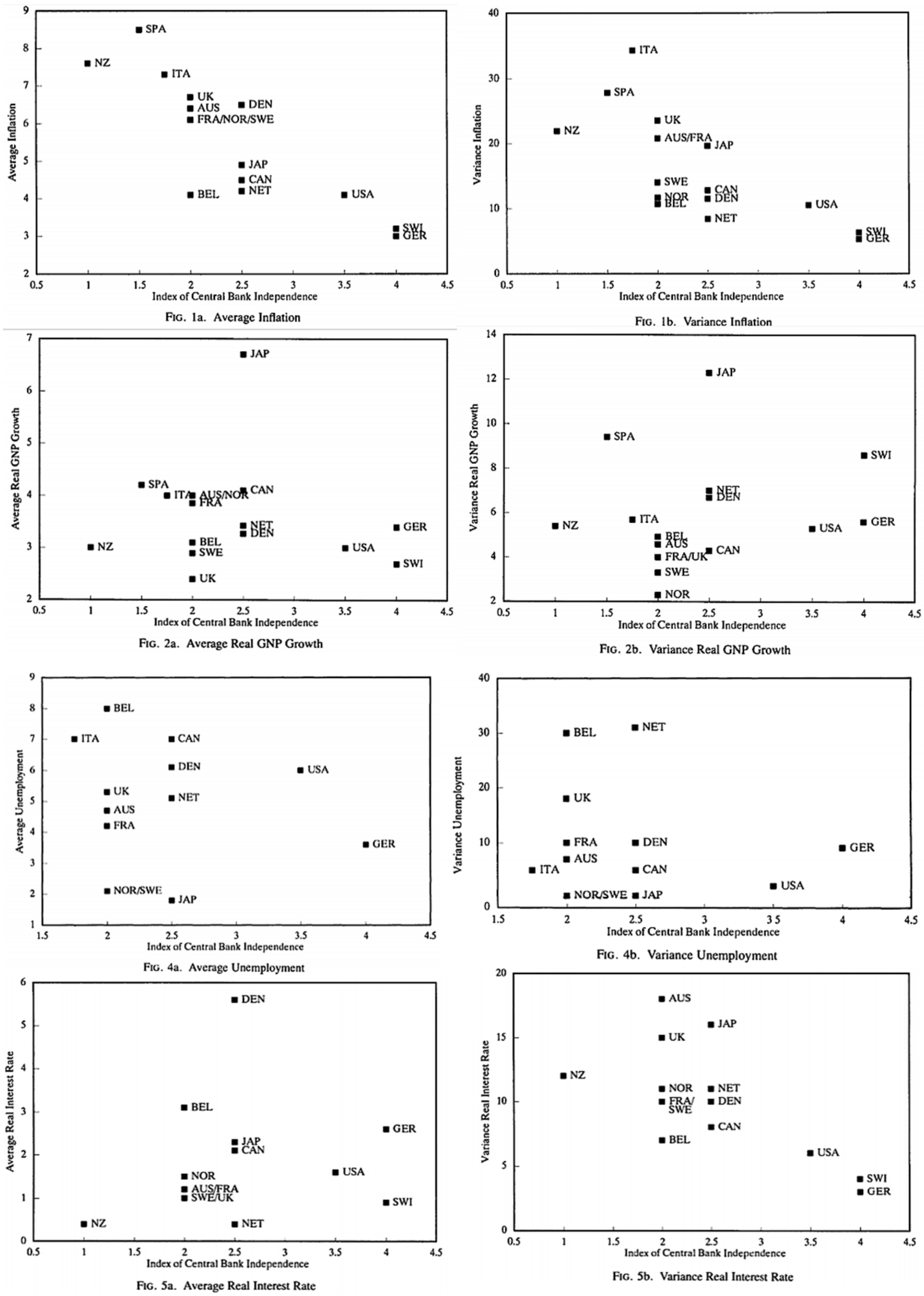


Figure 5: Graphical Representation of Alesina & Summers (1993) findings regarding the relationship of CBI with inflation, Gross National Product (GNP), unemployment and real interest rates.

According to the studies of Alesina & Summers (1993) and Arnone & Romelli (2013), the results signify a negative correlation between the CBI and the variability level of inflation. However, evidence suggests that CBI imposes no significant benefits to general macroeconomic performance of a state as no clear association is observed with output growth. Additionally, no relationship is denoted between CBI and either real interest rates nor unemployment. Overall, CBI in the wider economy has increased over the years and during these times of uncertainty with the word's economy collapsing under the Coronavirus pandemic, the notion CBI has been strongly affected since CBs became heavily involved in the economic response to the underlying Covid crisis.

word count: 2,500

References

Alesina, A. & Summers, L. H. (1993), 'Central bank independence and macroeconomic performance: some comparative evidence', *Journal of Money, credit and Banking* **25**(2), 151–162.

Arnone, M. & Romelli, D. (2013), 'Dynamic central bank independence indices and inflation rate: A new empirical exploration', *Journal of Financial Stability* **9**(3), 385–398.

Aschheim, J. (1973), 'Neutral money reconsidered', *PSL Quarterly Review* **26**(105).

Barro, R. J. & Gordon, D. B. (1983), 'Rules, discretion and reputation in a model of monetary policy', *Journal of monetary economics* **12**(1), 101–121.

Baumeister, C. & Benati, L. (2010), 'Unconventional monetary policy and the great recession-estimating the impact of a compression in the yield spread at the zero lower bound'.

Bernanke, B., Reinhart, V. & Sack, B. (2004), 'Monetary policy alternatives at the zero bound: An empirical assessment', *Brookings papers on economic activity* **2004**(2), 1–100.

Bernheim, B. D. (1987), 'Ricardian equivalence: An evaluation of theory and evidence', *NBER macroeconomics annual* **2**, 263–304.

Borio, C. & Zabai, A. (2016), Unconventional monetary policies: a re-appraisal, BIS Working Papers 570, Bank for International Settlements.
URL: *https://ideas.repec.org/p/bis/biswps/570.html*

Breedon, F., Chadha, J. S. & Waters, A. (2012), 'The financial market impact of uk quantitative easing', *Oxford Review of Economic Policy* **28**(4), 702–728.

Butt, N., Churm, R., McMahon, M. F., Morotz, A. & Schanz, J. F. (2014), 'Qe and the bank lending channel in the united kingdom'.

Cukierman, A. & Gerlach, S. (2003), 'The inflation bias revisited: theory and some international evidence', *The Manchester School* **71**(5), 541–565.

Disyatat, P. (2011), 'The bank lending channel revisited', *Journal of money, Credit and Banking* **43**(4), 711–734.

Glick, R. & Leduc, S. (2012), 'Central bank announcements of asset purchases and the impact on global financial and commodity markets', *Journal of International Money and Finance* **31**(8), 2078–2101.

Haldane, A., Roberts-Sklar, M., Young, C. & Wieladek, T. (2016), 'Qe: the story so far'.

Hamilton, J. D. & Wu, J. C. (2012), 'The effectiveness of alternative monetary policy tools in a zero lower bound environment', *Journal of Money, Credit and Banking* **44**, 3–46.

Hoshi, T. & Kashyap, A. (1999), 'The japanese banking crisis: Where did it come from and how will it end?', *NBER macroeconomics annual* **14**, 129–201.

Joyce, M., Lasaosa, A., Stevens, I. & Tong, M. (2010), 'The financial market impact of quantitative easing'.

Joyce, M., Miles, D., Scott, A. & Vayanos, D. (2012), 'Quantitative easing and unconventional monetary policy–an introduction', *The Economic Journal* **122**(564), F271–F288.

Joyce, M., Tong, M. & Woods, R. (2011), 'The united kingdom's quantitative easing policy: design, operation and impact', *Bank of England Quarterly Bulletin* .

Kapetanios, G., Mumtaz, H., Stevens, I. & Theodoridis, K. (2012), 'Assessing the economy-wide effects of quantitative easing', *The Economic Journal* **122**(564), F316–F347.

Martin, C. (2020), 'International monetary policy and institutions (es50078), topic 2. quantitative easing', `https://moodle.bath.ac.uk/mod/folder/view.php?id=818173`.

Martin, C. & Milas, C. (2012*a*), 'Quantitative easing: a sceptical survey', *Oxford Review of Economic Policy* **28**(4), 750–764.

Martin, C. & Milas, C. (2012*b*), 'A very large gamble: evidence on quantitative easing in the us and uk', *Institute for Policy Research. http://www. bath. ac. uk/ipr/policy-briefs/quantitative-easing. html* .

McLeay, M., Radia, A. & Thomas, R. (2014), 'Money creation in the modern economy', *Bank of England Quarterly Bulletin* p. Q1.

Meaning, J. & Zhu, F. (2011), 'The impact of recent central bank asset purchase programmes', *BIS Quarterly Review, December* .

Meier, A. (2009), *Panacea, curse, or nonevent? Unconventional monetary policy in the United Kingdom*, number 9-163, International Monetary Fund.

Rogoff, K. (1985), 'The optimal degree of commitment to an intermediate monetary target', *The quarterly journal of economics* **100**(4), 1169–1189.

Shah, I. & Malki, I. (2018), 'The portfolio balance channel: an analysis on the impact of quantitative easing on the us stock markets'.

Woodford, M. (2012), 'Methods of policy accommodation at the interest-rate lower bound'.

Yeyati, E. L., Ghezzi, P. & Broda, C. (2009), '1 the new global balance–part ii: Higher rates rather than weaker dollar in 2010', *Quantitative Easing* p. 21.

Appendices

Appendix 1 - QE and Interest Rates (Martin (2020)).

Considering two adjacent time periods t and $t+1$, and the specific interest rates:

i i_{2t}: risky, long-term private sector interest rate.
ii i_{2t}^b: risk-free, long-term government bond rate.
iii i_p: risk-free, short-term policy rate.

The interest rate on private sector borrowing from commercial often required a risk premium ρ_t as the loans are considered as risky asset for commercial banks.

$$i_{2t} = i_{2t}^b + \rho_t \tag{46}$$

Assuming that policy rate and government bond investments are perfect substitutes, employing the bond rate and policy rate relation, the return on an initial investment of £1 from a two period government bond is expressed as:

$$\textit{Return 1:}\;\; (1+i_{2t}^b)^2 - 1 = 1 + 2i_{2t}^b + (i_{2t}^b)^2 - 1 = 2i_{2t}^b + \underbrace{(i_{2t}^b)^2}_{\approx 0} \tag{47}$$

Additionally, the one period return of the same investment in the one period asset at the policy rate is expressed as:

$$\textit{Return 2:}\;\; (1+i_t^p)(1+E_t[i_{t+1}^p]) - 1 = i_t^p + E_t[i_{t+1}^p + \underbrace{i_t^p E_t[i_{t+1}^p]}_{\approx 0} \tag{48}$$

$$= i_t^p + E_t[i_{t+1}^p \tag{49}$$

Accounting for the perfect substitutability condition where *Return 1* = *Return 2*:

$$2i_{2t}^b = i_t^p + E_t[i_{t+1}^p] \Longrightarrow i_{2t}^b = \frac{i_t^p + E_t[i_{t+1}^p]}{2} \tag{50}$$

One can see that the bond interest rate equals the average of short term policy rates. Therefore, according to the underlying economic mechanism, with the reduction of policy rates by the Central Bank and under the assumption that market agents' expect the same policy action in the following time period then:

$$i_{2t}^b > \frac{i_t^p + E_t[i_{t+1}^p]}{2} \tag{51}$$

Consequently, demand of bonds goes up driving i_{2tp} down until:

$$i_{2t}^b = \frac{i_t^p + E_t[i_{t+1}^p]}{2} \tag{52}$$

However, in the case where the two assets are not identical and government bonds are less attractive due to being more risky and less liquid, investors would rather substitute their portfolios with shorter duration assets. Thus, capturing the underlying effect a premium is introduced on the interest rates such that:

$$i_{2t}^b \;\;=\;\; \frac{i_t^p + E_t[i_{t+1}^p]}{2} \;\;+\;\; \underbrace{\rho_t^2}_{\textit{liquidity premium}} \tag{53}$$

Thus a bond risk premium is included in such that:

$$i_t = \frac{i_t^p + E_t[i_{t+1}^p]}{2} + \underbrace{\rho_t^2}_{liquidity\ premium} + \underbrace{\rho_t}_{bond\ premium} \tag{54}$$

where i_t represents the policy tool of conventional monetary policy and it is adjusted via the reduction risk premia ρ_t and ρ_t^2, altering the private sector's interest rate i_p and stimulating aggregate demand.

Appendix 2: Broad Money Derivation (Martin (2020)).

According to the Basel regulations, commercial banks are required to hold sight-deposit accounts in the Central Bank, with a minimum amount of their net worth maintained as a cushion against the risk of loan devaluation, known as minimal capital adequacy. The bank reserves are given as the summation of those bank deposits at the Central Bank and the cash holdings in the bank vault. Additionally, the money base M_0 is defined as the total currency in circulation CU plus the reserves R of commercial banks given as a fraction of their deposits D, thus:

$$\underbrace{R}_{Reserves\ Lower\ Bound} \geq \underbrace{rrr}_{Legal\ Minimum\ Reserve\ Ratio} \times D \tag{55}$$

However, in most cases actual reserves exceed the legal minimum level such that:

$$\underbrace{R}_{Actual\ Reserves} = \underbrace{rr}_{Actual\ Reserve\ Ratio} \times D, \quad where\ \ rr \geq rrr \tag{56}$$

The relation between broad money M_1 determined by the public, and narrow money M_0 solely determined by the Central Bank is expressed by the money multiplier m, where by definition:

$$m = \frac{M_1}{M_0} = \frac{CU + R}{CU + D} \tag{57}$$

Expressing Reserves R as a fraction of Deposits D, $R = rrD$ and accounting for the fact that the private sector holds a fraction of Broad money M_1 in currency CU such that $CU = ccM_1$, then M_0 and M_1 are expressed as:

$$M_0 = ccM_1 + rrD \tag{58}$$

$$M_1 = ccM_1 + D \tag{59}$$

$$\longrightarrow D = (1 - cc)M_1 \tag{60}$$

Dividing Equation (59) with (58):

$$m = \frac{M_1}{M_0} = \frac{ccM_1 + D}{ccM_1 + rrD}. \tag{61}$$

Substituting the expression $D = (1 - cc)M_1$:

$$\Longrightarrow m = \frac{ccM_1 + M_1 - ccM_1}{ccM_1 + rr(M_1 - ccM_1)} = \frac{M_1(cc + 1 - cc)}{M_1(cc + rr - ccrr)} = \frac{1}{cc + rr(1 - cc)} \tag{62}$$

Appendix 3: Commercial Banks Profitability (Martin (2020)).

Expressing the interest rate on lending as i^L and the interest rate on reserves as i^R, and assuming that the hare of reserves on total assets is denoted as ϕ^R, then under the assumption of no other assets included on commercial banks balance-sheet, other than loand and reserves, then the return on assets i^A of a commercial bank is given by:

$$i^A = \phi^R i^R + (1 - \phi^R) i^L \tag{63}$$

Accounting for the fact that interest rate on reserves is less than the interest rate on loads, since they are seen as riskier assets, then:

$$i^R < i^L \ , \tag{64}$$

then by increasing the reserves to loans ratio, the weighing on i^R denoted by ϕ^R results to a decrease in i^A expressing the weighted average between returns on reserves and returns on loans. Therefore, the total profitability of the commercial bank declines, motivating the rebalancing of its pre-LASP portfolio composition by increasing lending leading to the increase of bank deposits.

1. Briefly discuss the impact of non-stationarity in multivariate regression analysis. In this context, develop the concept of cointegration, with a focus on the contribution of cointegration in multivariate analysis. Develop models for cointegrated variables in both a single equation and Vector Autoregression setting and illustrate methodologies for modeling, testing, and interpretation of such models. Give examples of applications where such models may be used. Provide some intuition for the results of your models and their interpretation.

The interference between applied econometrics and the underlying economic theory appears as an uneasy compromise where a tenuous theoretical base is employed for the pragmatic justification of various acknowledged procedures. Hylleberg & Mizon (1989), emphasise that despite the recognised non-stationary nature of various economic time series, it was not until the 1970s, when economic modelling was impacted by time series analysis, and more attention was attributed to the features of the series. Additionally, Granger et al. (1974), underlines the frequently observed phenomenon in econometric literature of time series regressions characterised by a high coefficient of multiple correlation R^2 combined with a remarkably low Durbin-Watson statistic, suggesting the spurious nature of the economic data arising from the inappropriate formulation of the errors autocorrelation structure.

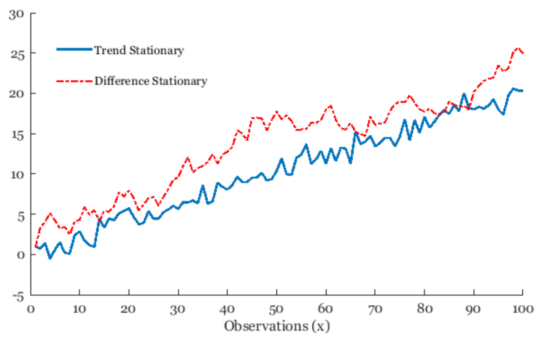
In determining the implications of non-stationarity in multivariate regression analysis, the concept of stationarity and the major forms of non-stationarity are introduced. Employing the mathematical analysis of Hylleberg & Mizon (1989), a general vector N, composed of random variables X_t, indicating the observation of a time series is characterised as covariance stationary under the constraints of time invariant mean vector and covariance matrices. Additionally, the primary non-stationarity forms include the seasonal and trend fluctuations which can be separated into two categories - stochastic and deterministic - enabling x_t to be demonstrated as a factor of five components.

$$X_t = T_t + S_t + \mu_t + \xi_t + \epsilon_t \quad , \tag{1}$$

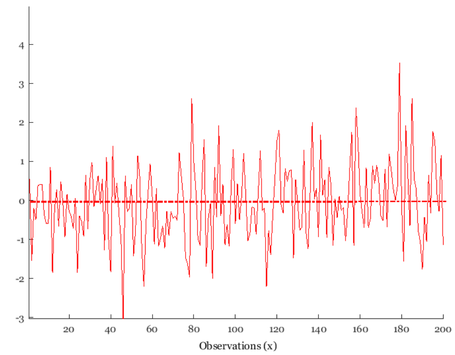
where T_t is a polynomial in t signifying deterministic trends (Trend Stationary), S_t denotes all deterministic seasonals, μ_t represents possible stochastic trends including integrated processes and random walks (Difference Stationary), ξ_t denotes stochastic seasonal processes and finally ϵ_t denotes an independently and identically distributed random vector named white noise illustrated by Figure 2. According to standard econometric analysis, the orthogonality condition is required to be satisfied between all components, else the model is mis-specified, giving rise to omitted variable bias, misleading parameter interference, and the non-stationary behaviour of the disturbance term. Additional dangers inherent in including non-stationary variables in multiple regression analysis according Sims (1988), involve the presence of amplifying shock propagation in the system and the prevention of the respective asymptotic analysis, namely t-statistics and F-statistics to follow their distinctive distributions. Consequently, this leads to an overestimated relationship between the dependent and independent variables, the invalidity of hypothesis tests about the regression parameters and sub-optimal forecasts.

Wold's theorem states a stationary time series which includes no deterministic components Y_t is represented by an infinite order moving average, often approximated by a finite order autoregressive moving average process (Engle & Granger (1987)).

$$Y_t - \phi_1 Y_{t-1} - ... - \phi_p Y_{t-p} = \alpha_t - \theta_1 \alpha_{t-1} - ... - \theta_q \alpha_{t-q} \quad , \tag{2}$$



(a) Figure 1. Graphical Representation of Trend Stationary and Difference Stationary series.



(b) Figure 2. Graphical Representation of white noise process.

where α_t is a white noise process. Rearranging and expressing the equation in a polynomial form,

$$\phi(B)Y_t = \theta(B)\alpha_t \quad , \tag{3}$$

where $\phi(B)$ and $\theta(B)$ represent the polynomial lag operators with adequate roots to secure the uniqueness of representation and stationarity of Y_t. However, in the case where a time series X_t, experiences stochastic non-stationarity, then it requires differencing a sufficient number of times to be reduced to stationarity. Thus, an integer d exist such that

$$\nabla^d X_t = Y_t \quad . \tag{4}$$

Finally, substituting Equation (3) into Equation (2), the time series X_t, are expressed as an autoregressive integrated moving average process (A.R.I.M.A) of order (p, d, q).

$$\phi(B)\nabla^d X_T = \theta(b)\alpha_t \quad . \tag{5}$$

By definition, a series with no deterministic component, represented by an invertable ARMA model after being differenced d times is known to be integrated of order d, signified as $X_t \sim I(d)$. A substantial difference between a stationary $I(0)$ and non-stationary $I(d)$ time series includes the theoretical infinite variance of the $I(1)$ series resulting from the superposition of low frequencies on the model's long-run component. Consequently, the magnitude difference between the variances and the dominant long swings of $I(d)$ relative to $I(0)$, it is fundamentally true that any linear combination Z_t will result to be of order $I(d)$.

$$Z_t = X_t - \alpha Y_t \sim -I(d) \tag{6}$$

In the case where $Z_t \sim I(d-b)$, $b > 0$ an exceptional constraint applies to the long-run elements of the investigated series. Examining the case where the series are of the same integration order ,$d = b$, the resulting linear combination appears to be $\sim I(0)$, and does not suffer from low frequencies. Hence, variable α, indicates that the vast majority of the long run elements of both X_t and Y_t cancel out, representing a common stochastic trend between the two series appearing to be bounded under the same co-movement. Put differently, the fact that the two series cannot drift too far apart implies a special relationship reflected by the underlying statement of their differences being stationary. Finally, the formal definition of the aforementioned idea, originally developed by Granger & Weiss (1983) is:

The elements of the vector X_t, are defined as cointegrated of order d, b, signified as $X_t \sim CI(d, b)$ under the following specifications.
(i) all components of X_t are of order $I(d)$;
(ii) a cointegrating vector α ($\neq 0$) exists, such that $Z_t = a'X_t \sim I(d - b), b > 0$.

In the case where $d = b = 1$, the intuition behind cointegration suggests that if all elements of X_t are $I(1)$, then the resulting equilibrium error would be stationary $I(0)$, giving rise to equilibrium since Z_t would often cross the zero axis without significantly deviating. Alternatively, if X_t does not exhibit cointegration, then the dynamics of Z_t, would be characterised by an unbounded drift where zero-crossings rarely occur. Additionally, in the case where all the components of X_t are already stationary, then the equilibrium error Z_t does not have idiosyncratic properties.

Granger & Weiss (1983) also distinguish between the two scenarios namely a bivariate system where X_t is composed of $N = 2$ elements, and the multivariate system with $N > 2$ elements. In the former, the sufficient cointegration condition suggests that the coherence between the time series at zero frequency equals one, whilst in the latter the possibility of various equilibrium relations that specify the joint behaviour of the variables arises, resulting in more than one cointegrating vectors α. Assuming the number of unique linearly independent cointegrated relationships between the variables is denoted by r, with $r \leq N - 1$, all linear relationships are assembled in the $N \times r$ array, α. Hence, the rank of α equals to r, and is known as "cointegration rank" of vector X_t.

One of the major attractions of the investigated concept, includes the development of a framework that enables the modelling of short-run dynamics and the long-run equilibrium, $z_t = \alpha' X_t$, which governs information associated with economic theory providing a meaningful analysis of $I(1)$ series. Even though the statistical problem of spurious regressions can be eliminated by differencing the time series, according to Hendry & Mizon (1978), the appropriateness of the aforementioned model specification requires testing rather than presumption as it leads to the detriment of valuable long-run information. Specifically, in the case where X_t and Y_t are both $I(1)$, the underlying model would estimate

$$\Delta Y_t = \beta \Delta X_t + \epsilon_t \quad . \tag{7}$$

Considering the long run definition where $Y_t = Y_{t-1}$ and $X_t = X_{t-1}$, all the difference terms collapse to zero, $\Delta Y_t = \Delta X_t = 0$, and no meaningful long-run relationship is obtained.

The idea behind error correction mechanisms as per Sargan (1964), is that a fraction of disequilibrium arising in one period is accounted and corrected in the following period. Engaging the Granger Representation Theorem, any cointegrating relation can be reduced to an Equilibrium Correction Model (ECM). Typically, in the bivariate regime, the error correction representations associate the change in one variable both to past changes in equilibrium variables and equilibrium errors, providing a statistical way of accounting for long-run equilibria. For example assuming that

$$x_t = x_{t-1} + e_t \sim I(1) \tag{8}$$

$$y_t = \gamma x_t + u_t \sim I(1) \implies u_t = y_t - \gamma x_t \tag{9}$$

In the case where $u_t \sim I(0)$, then $y_t - \gamma x_t \sim I(0)$, whereas the associated cointegrating vector is denoted as $= [1, -\gamma]$. Subtracting, y_{t-1} from both Equation (8) and substituting equation (9),

$$y_t = \gamma x_t + u_t \tag{10}$$

$$\Delta y_t = -y_{t-1} + \gamma x_t + u_t = -y_{t-1} + \gamma\left(x_{t-1} + e_t\right) + u_t$$

$$= -\underbrace{\left(y_{t-1} - \gamma x_{t-1}\right)}_{Level\ Equation} + \underbrace{\gamma e_t + u_t}_{\nu_t}$$

where, $\nu_t \sim I(0)$ by construction. Therefore, the error correction model, employs both level terms and first differences of the variables as illustrated below.

$$\Delta y_t \;=\; \beta_1 \Delta x_t \;+ \beta_2 \underbrace{\left(y_{t-1} - \gamma x_{t-1}\right)}_{Error\ Correction\ Term} \;+\; u_t \tag{11}$$

According to Equation (11), γ changes between adjacent time periods as a result of changes in the explanatory variable, correcting for any disequilibrium that arose, determining the long-run relation between y and x. Similarly, β_1 defines the system's short-run dynamic adjustment mechanism, relative to the short-term relationship between changes in the variables, and β_2 signifies the rate of convergence of the system towards the long-run equilibrium, also known as speed of adjustment. To ensure the equilibrium mechanism is consistent with the long-run relation, the constraint of $\beta_2 < 0$ is necessary so that any proportion of disequilibrium that arose in a particular time period is corrected in the following period by driving the system back to the long-run equilibrium state (Salmon (1982)). Finally, the standard procedure of statistical interference and OLS regressions can be employed for estimating β_2 as the error correction term is of order $I(0)$.

Generalising to the n-variable case, assuming that all variables are $I(1)$, the system is specified by a VAR model of order p, where each endogenous variable is a linear function of p lags of itself and the remaining $n-1$ variables:

$$\underbrace{X_t}_{n\times 1} = \underbrace{A_1}_{n\times n}\underbrace{X_{t-1}}_{n\times 1} + \underbrace{A_2}_{n\times n}\underbrace{X_{t-2}}_{n\times 1} + \ldots.. + \underbrace{A_p}_{n\times n}\underbrace{X_{t-p}}_{n\times 1} + \underbrace{\epsilon_t}_{n\times 1} \quad , \tag{12}$$

where ϵ_t signifies a vector of stationary errors, and the formal Vector Error Correction Model (VECM) representation becomes:

$$\Delta X_t = + \underbrace{\Pi X_{t-1}}_{error\ correction\ term} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \;\epsilon_t \tag{13}$$

where,

$$\Pi = \sum_{i=1}^{p} A_i - I \quad ; \quad \Gamma_i = \sum_{j=i+1}^{p} A_i \tag{14}$$

To support the above argument, a VAR(3) system is employied and the parametric representation of the system into a VECM is derived (see Appendix 1).

According to literature, there are numerous approaches for the estimation of cointegration relations namely, the Augmented Least Squares (Park (1992)), Spectral Regression (Phillips et al. (1988)), Principal Components (Stock & Watson (2002)), Maximum Likelihood (Johansen & Juselius (1990)), Modified Box-Tiao (Bewley et al. (1994)). The current discussion however, expands on two of the most commonly used namely, the Engle- Granger and Johansen approaches. The main objective includes the examination of whether an equilibrium relation between two $I(1)$ variables, x_t and y_t exists. Engle & Granger (1987) methodology examines whether the variables are cointegrated of order $CI(1,1)$, which is broken down to three main steps.

Step 1 includes identifying the integration order of all variables to ensure they are of the same order of integration. Therefore, various unit root tests are carried out, namely Dickey-Fuller (DF), the augmented Dickey-Fuller (ADF) or the Phillips-Perron test (PP). The second step includes the estimation of the long run relationship between the two variables using the following OLS regression:

$$y_t = \gamma x_t + \epsilon_t \ , \tag{15}$$

from which the residuals $\hat{\epsilon}$ are saved. Under the assumption that a cointegrating relationship between the variables exist, the OLS regression yields to be a super-consistent estimator of the regressed parameters as the deviations from the long-run equilibrium $\hat{\epsilon}$ are of order $I(0)$. The specification of the residuals integration order is determined by running a DF test where the critical values of Engle and Granger are employed.

$$\Delta\hat{\epsilon} = \delta\hat{\epsilon}_{t-1} + \nu_t \ \ , where \ \nu_t \sim iid. \tag{16}$$

if the null hypothesis of $\delta = 0$, is not rejected then the residuals contain a unit root and one concludes the examined variables are not cointegrated. Alternatively, if the null is rejected, the residuals are stationary and the variables are cointegrated of order $N(1,1)$. In the case where the residuals do not appear as white noise, an alternative approach would be to run an ADF test to determine whether $\hat{\epsilon}$ suffer from serial correlations:

$$\Delta\hat{\epsilon} = \gamma\hat{\epsilon}_{t-1} + \sum_{i=1}^{n} {}_i\Delta\hat{\epsilon}_{t-i} + \nu_t \tag{17}$$

Similarly, in the case where $-2 < \gamma < 0$, the stationarity of $hat\epsilon$ is established and variables are $CI(1,1)$. The final step of the process includes the construction and estimation of the error-correction model using $\hat{\epsilon}$. Following the first two steps, the error correction form of the variables is:

$$\Delta y_t = \beta_1\Delta x_t + \beta_2\hat{\epsilon}_{t-1} + \nu_t \tag{18}$$

where, $\hat{\epsilon}_{t-1} = y_{t-1} - \hat{\gamma}x_{t-1}$ and the cointegration vector is given by $[1 - \gamma]$. Consequently, all parameters are stationary so inferences are valid. However, the investigated method suffers from various limitations. Firstly, the finite sample problem expressed by the lack of power in both unit root and cointegration tests arises, which can be diminished with an increasing sample size. Additionally, the singe equation approach gives rise to simultaneous bias due to the asymmetric treatment of the variables. This can be treated by repeating the process and normalising the remaining variables, ensuring that the specification order has no impact on the results. Furthermore, performance of hypothesis testing on the long-run system relation $\hat{\gamma}$, is not feasible. Finally, the most significant problem regarding the residual based technique involves its weakness to identify the most optimal long-run relation in a multivariate system.

A common approach that deals with the determination of all possible r relationships and allows their estimation is Johansen's methodology. Recalling the VECM representation represented by Equations (13) and (14), and investigating the case where $rank(\Pi) = r$, $0 < r < n-1$, the matrix Π is decomposed to two alternative $n \times r$ matrices:

$$\Pi = \alpha\beta' \tag{19}$$

where α and β include all r adjustment vectors and cointegrating vectors respectively. Hence, in the case where $X_{t-1} \sim I(1)$, then by definition $\Pi X_{t-1} = \alpha(\beta' X_{t-1})$.

$$\Pi X_{t-1} = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \cdots \\ \cdots \\ \alpha_{n-1} \\ \alpha_n \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \cdots & \beta_{n-1} & \beta_n \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \\ \cdots \\ \cdots \\ X_{n-1_{t-1}} \\ X_{nt-1} \end{pmatrix} \tag{20}$$

The decomposition of Π does not have a unique solution. For any invertible matrix K, various linear transformations of β exist such that $\beta^{*\prime} = K\beta'$. Thus, the cointegrated vector does not have a unique representation, giving rise to the long-run identification problem. Unlike the Engle Granger approach, in order derive an exact identification of the investigated system, it necessities r amount of restrictions, either of statistical or empirical nature, to be imposed evenly across each individual long-run relation, summing to r^2 restrictions (for decomposition process see Appendix 1). Imposing the equilibrium condition, setting $\Delta X = 0$, the long-run representation reduces to:

$$\Pi X^* = 0 \Longrightarrow \alpha(\beta' X^*) = 0 \Longrightarrow \beta' X^* \sim I(0) = 0. \tag{21}$$

In the interesting case where $0 < Rank(\Pi) = Rank() = (\beta') < n$, the total number of non-zero solutions of the system equals $\beta' X$. The outlined decomposition process represents the error correction mechanism of the system $\alpha\beta' X_{t-1}$, ensuring its periodical convergence towards the long-run relation β. Due to the reverting nature of the process, any imbalances are reflected in the subsequent periods by deviations of $\beta' X_{t-1} = \xi_{t-1} \neq 0$, causing ΔX to respond accordingly, driving the system back to equilibrium. Thus, $\alpha \times \beta$ accounts for the magnitude and direction of the disturbance, requiring the two terms to be of opposite signs.

The cointegration test between the system's variables $X's$, is determined by the $Rank(\Pi)$, via its eigenvalues λ_i, where

$$\lambda_1 \geq \lambda_2 \geq ..... \geq \lambda_k \tag{22}$$

Particularly, the rank of the matrix is defined as the number of non-zero characteristic roots. In the case where $Rank(\Pi)$, is not significantly different from zero, $\lambda_i = 0, \forall i$, then the variables are not cointegrated and $\ln(1-\lambda_i) = 0$. Similarly, if the system is stationary, and all unit roots lie within the unit circle, $0 \leq \lambda_i < 1, \forall i$ then $\ln(1-\lambda_i) < 0$. Finally, if $Rank(\Pi) = 1$, then $\ln(1-\lambda_1)$ would be negative and $\ln(1-\lambda_i) < 0, \ \forall i > 1$. There are two sequential Johansen tests whose statistics rely on the outlined idea, first of which is the joint trace test:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1-\hat{\lambda}_i), \tag{23}$$

where T denotes the number of observations and $\hat{\lambda_i}$ signifies the calculated value of the i^{th} eigenvalue form Π. The null hypothesis assumes the number of cointegrating vectors in the system is less or equal to r. The second cointegration test statistic includes the maximum eigenvalue test formulated as:

$$\lambda_{max}(r, r+1) = -Tln(1 - \hat{\lambda_{r+1}}) \tag{24}$$

which unlike λ_{trace}, each eigenvalue is tested separately under the null hypothesis of r cointegrating vectors in the system, against the alternative of $(r+1)$ vectors. Both statistics' distributions are non-standard and their critical values, estimated by Johansen & Juselius (1990), depend on the regression form deterministic terms in the VECM (See Appendix 2), and the value of non stationary components, $(k-r)$. One repeats the testing sequence, increasing value of r, as demonstrated by Table 1, until the test statistic is less that the critical value and the null hypothesis is no longer rejected, concluding the total number of cointegrating vectors in the system.

Null	$\lambda_{trace} - Alternative$	$\lambda_{max} - Alternative$
$H_0 : r = 0$	H1: $0 <$ r $\leq n$	H1:r=1
$H_0 : r = 1$	H1: $1 <$ r $\leq n$	H1: r=2
$H_0 : r = 2$	H1: $2 <$ r $\leq n$	H1: r=3
...	...	...
$H_0 : r = n-1$	H1: r $<$ r $\leq n$	H1: r=n

Table 1: Testing sequence, outlining null hypotheses and alternative hypotheses for both λ_{trace} *and* λ_{max}.

If the model specifications contain more complex restrictions, the normalisation condition would fail to satisfy them, having no impact on the cointegration vector. One way to determine whether a restriction is binding is the Johansen Hypothesis testing that uses the characteristic roots of the restricted model λ_i^* with r non-zero characteristic roots and m restrictions and the unrestricted model λ_i. The test statistic of the hypothesis is:

$$-T\sum_{i=1}^{r}[ln(1-\lambda_i) - ln(1-\lambda_i^*)] \sim \chi^2(m) \tag{25}$$

In the case where the null hypothesis rejects, the restrictions are not binding resulting to a big difference between the models, therefore the unrestricted model is employed. If however, the null does not reject then the differences between the two models are insignificant and the restricted model is employed since is more parsimonious.

The discussed methodology of ECM and VECM model analysed, has a wide range of real life applications including the study of Mills and Mills 1991 investigating the phenomenon of cointegration in four international markets namely US, UK, Germany and Japan and their associated dynamics, by employing daily closing observation for the redemption yield. Additional papers use the examined models to estimate the equilibrium long-run exchange rate known as purchasing power parity between two or more currencies identifying ant potential Granger causality between variables as well as the long-run dynamics of various macroeconomic parameters.

References

Bewley, R., Orden, D., Yang, M. & Fisher, L. A. (1994), 'Comparison of box—tiao and johansen canonical estimators of cointegrating vectors in vec (1) models', *Journal of Econometrics* **64**(1-2), 3–27.

Engle, R. F. & Granger, C. W. (1987), 'Co-integration and error correction: representation, estimation, and testing', *Econometrica: journal of the Econometric Society* pp. 251–276.

Granger, C. W., Newbold, P. & Econom, J. (1974), 'Spurious regressions in econometrics', *Baltagi, Badi H. A Companion of Theoretical Econometrics* pp. 557–61.

Granger, C. W. & Weiss, A. A. (1983), Time series analysis of error-correction models, *in* 'Studies in econometrics, time series, and multivariate statistics', Elsevier, pp. 255–278.

Hendry, D. & Mizon, G. (1978), 'Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the bank of england', *Economic Journal* **88**(351), 549–63.
URL: *https://EconPapers.repec.org/RePEc:ecj:econjl:v:88:y:1978:i:351:p:549-63*

Hylleberg, S. & Mizon, G. E. (1989), 'Cointegration and error correction mechanisms', *The Economic Journal* **99**(395), 113–125.

Johansen, S. & Juselius, K. (1990), 'Maximum likelihood estimation and inference on cointegration—with applications to the demand for money', *Oxford Bulletin of Economics and statistics* **52**(2), 169–210.

Park, J. Y. (1992), 'Canonical cointegrating regressions', *Econometrica: Journal of the Econometric Society* pp. 119–143.

Phillips, P. C. et al. (1988), *Spectral regression for cointegrated time series*, Cowles Foundation for Research in Economics at Yales University.

Salmon, M. (1982), 'Error correction mechanisms', *The economic journal* **92**(367), 615–629.

Sargan, J. D. (1964), 'Wages and prices in the uk', *Econometric analysis for national economic planning* pp. 29–59.

Sims, C. A. (1988), 'Bayesian skepticism on unit root econometrics', *Journal of Economic dynamics and Control* **12**(2-3), 463–474.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American statistical association* **97**(460), 1167–1179.

Appendices

Appendix 1 – From VAR to VECM

$$\underbrace{X_t}_{n\times 1} = \underbrace{A_1}_{n\times n}\underbrace{X_{t-1}}_{n\times 1} + \underbrace{A_2}_{n\times n}\underbrace{X_{t-2}}_{n\times 1} + A_{3n\times n}\underbrace{X_{t-3}}_{n\times 1} + \underbrace{\epsilon_t}_{n\times 1} \tag{26}$$

Step 1: Subtracting X_{t-1} from both sides:

$$\Delta X_t = (A_1 - I)X_{t-1} + A_2 X_{t-2} + A_3 X_{t-3} + \epsilon_t \tag{27}$$

Step 2: Identify the biggest lag X_{t-3}, add and subtract $A_3 X_{t-2}$ from the R.H.S:

$$\Delta X_t = (A_1 - I)X_{t-1} + A_2 X_{t-2} + A_3 X_{t-2} - A_3 X_{t-2} + A_3 X_{t-3} + \epsilon_t \tag{28}$$

$$= (A_1 - I)X_{t-1} + (A_2 + A_3)X_{t-2} - A_3 \Delta X_{t-2} + \epsilon_t$$

Step 3: Identifying the following greatest lag, X_{t-2}, add and subtract $(A_2 + A_3)X_{t-1}$ from the R.H.S:

$$\Delta X_t = (A_1 - I)X_{t-1} + (A_2 + A_3)X_{t-2} - (A_2 + A_3)X_{t-1} + (A_2 + A_3)X_{t-1} - A_3 \Delta X_{t-2} + \epsilon_t \tag{29}$$

$$= \underbrace{(A_1 + A_2 + A_3 - I)}_{\Pi} X_{t-1} - \underbrace{(A_2 + A_3)}_{\Gamma_1} \Delta X_{t-1} - \underbrace{A_3}_{\Gamma_2} \Delta X_{t-2} + \epsilon_t$$

The VECM represents a system of equations comparable to the Augmented Dickey Fuller test. Generally starting form a VAR(p), the derived VECM is of order (p-1). Imposing the equilibrium condition, where all ΔX_{t-i} collapse to zero, and the error terms converge to their expected value of zero, Π can be interpreted as the long-run coefficient matrix. Additionally, in the case where Π, has a full-rank $(r = n)$, then its inverse matrix Π^{-1} exists, which when multiplied with Equation (13) and re-arranged becomes:

$$\Pi^{-1}\Delta X_t = \Pi^{-1}\Pi X_{t-1} + \Pi^{-1}\sum_{i=1}^{p-1}\Gamma_i \Delta X_{t-i} + \Pi^{-1}\epsilon_t \tag{30}$$

$$\implies X_{t-1} = \Pi^{-1}\Delta X_t + \Pi^{-1}\sum_{i=1}^{p-1}\Gamma_i \Delta X_{t-i} + \Pi^{-1}\epsilon_t.$$

Observing the above specification, the R.H.S of the equation is a linear combination of stationary variables, resulting in X_t to be $I(0)$ eliminating the possibility for cointegration. On the contrary, if the $Rank(\Pi) = 0$, where Π equals the null matrix and all eigenvalues, $(\Pi) = 0$, no linear relationship between the $I(1)$ variables exists, and the variables are not cointegrated. Hence, the equation form reduces to a stationary $VAR(p-1)$ by taking differences:

$$\Delta X_t = \sum_{i=1}^{p-1}\Gamma_i \Delta X_{t-i} + \epsilon_t \tag{31}$$

Appendix 2 – VACM and error correction.

Employing a VECM representation with $k = 3$ variables and $r = 2$ cointegration equations form the above specification the system is expressed as:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \epsilon_t$$

Decomposing vector Π into $\alpha \times \beta'$, the matrix form is the system is reduces to:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \Delta y_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}$$

$$\Delta y_t = \alpha \, (k \times r) \quad \times \quad \beta' \, (r \times k) \, y_{t-1} + \Gamma_1 \, (k \times k) \, \Delta y_{t-1} + \epsilon_t$$

In order to estimate the system and retrieve estimates for all the parameters $r \times (k-1)$ restrictions should be imposed in the system. Employing the methodology imposed in E-views, the restrictions are imposed according to the identity matrix on the first $k-1$ columns of β'. Therefore,

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta 21 & \beta_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Longrightarrow \beta_{11} = \beta_{22} = 1 \quad and \quad \beta_{12} = \beta_{21} = 0$$

Expanding the matrix form into a system of equations the following relations are obtained.

$$\Delta y_{1t} = \alpha_{11} \left(\beta_{11} y_{1(t-1)} + \beta_{11} y_{1(t-1)} + \beta_{12} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right)$$

$$+\alpha_{12} \left(\beta_{21} y_{1(t-1)} + \beta_{22} y_{1(t-1)} + \beta_{23} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right) + \underbrace{\gamma_{11} \Delta y_{t-1} + \gamma_{12} \Delta y_{t-2} + \gamma_{13} \Delta y_{t-3}}_{Short\ term\ effects} + \epsilon_{1t}$$

$$\Delta y_{2t} = \alpha_{21} \left(\beta_{11} y_{1(t-1)} + \beta_{11} y_{1(t-1)} + \beta_{12} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right)$$

$$+\alpha_{22} \left(\beta_{21} y_{1(t-1)} + \beta_{22} y_{1(t-1)} + \beta_{23} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right) + \underbrace{\gamma_{21} \Delta y_{t-1} + \gamma_{22} \Delta y_{t-2} + \gamma_{23} \Delta y_{t-3}}_{Short\ term\ effects} + \epsilon_{2t}$$

$$\Delta y_{3t} = \alpha_{31} \left(\beta_{11} y_{1(t-1)} + \beta_{11} y_{1(t-1)} + \beta_{12} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right)$$

$$+\alpha_{32} \left(\beta_{21} y_{1(t-1)} + \beta_{22} y_{1(t-1)} + \beta_{23} y_{2(t-1)} + \beta_{13} y_{1(t-3)} \right) + \underbrace{\gamma_{31} \Delta y_{t-1} + \gamma_{32} \Delta y_{t-2} + \gamma_{33} \Delta y_{t-3}}_{Short\ term\ effects} + \epsilon_{3t}$$

Appendix 3 – 5 Cases of Deterministics of VECM

Employing the general expression for a VECM(p-1) model:

$$\Delta X_t = \alpha_0 + \alpha_1 t + \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$

the model is be divided into five different deterministic cases.

(I) No intercept and no trends, $\alpha_0 = 0$ and $\alpha_1 = 0$, where data have zero mean.

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$

(II) No linear trends in the series and restricted intercepts emerging in the cointegrating relations, $\alpha_0 = \Pi\mu, \alpha_1 = 0$.

$$\Delta X_t = \Pi(X_{t-1} + \ \mu) + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$

(III) Data suffering from stochastic trends and non zero intercepts in the cointegrating relations, $\alpha_0 \neq 0, \alpha_1 = 0$.

$$\Delta X_t = \alpha_0 + \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$

(IV) Restricted trends appearing in the cointegration relation and unrestricted intercepts, $\alpha_0 \neq 0, \alpha_1 = \Pi\gamma$.

$$\Delta X_t = \alpha_0 + \Pi[X_{t-1} + \gamma(t-1)] + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$

(V) Intercepts and trends are unrestricted, $\alpha_0 \neq 0, \alpha_1 \neq 0$.

$$\Delta X_t = \alpha_0 + \alpha_1 t + \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + u_t$$