UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Bayesian Econometrics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Parush Arora

Dissertation Committee:
Associate Professor Ivan Jeliazkov, Chair
Professor Fabio Milani
Associate Professor Yingying Lee

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Professor Ivan Jeliazkov for his incredible guidance and insight throughout the degree along with his exceptional mentorship as a research collaborator, instructor, and supervisor.

Yingying Lee and Fabio Milani, and all other professors within the Economics department at UC Irvine who at various points offered me guidance, critiques, or conversations that sparked academic inquiry. I especially thank them for the excellent instruction I received through the years, which built the foundations from which I could explore my interests in Economics.

I am grateful to my family for backing and supporting me throughout all these years. I would also like to thank my friends who kept me sane. I cherish the experiences we had at UC Irvine.

VITA

Parush Arora

EDUCATION

Doctor of Philosophy in Economics **2024**
University of California, Irvine *Irvine, California*

Master of Arts in Economics **2018**
Madras School of Economics *Chennai, India*

Bachelor of Arts in Economics **2016**
Hansraj College, University of Delhi *New Delhi, India*

RESEARCH EXPERIENCE

Graduate Research Assistant **2022**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2019–2024**
University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

Essays in Bayesian Econometrics

By

Parush Arora

Doctor of Philosophy in Economics

University of California, Irvine, 2024

Associate Professor Ivan Jeliazkov, Chair

This dissertation is comprised of three chapters.

Chapter 1 is a reprint of my job market. The chapter considers the efficient estimation of opinion pools in the Bayesian paradigm and extends their application to cases where the number of competing models exceeds the number of observations. An appropriate Bayesian formulation and estimation algorithm is proposed which 1) accommodates any proper scoring rule and 2) allows the weights to shrink towards any possible combination. This flexibility makes the Bayesian opinion pool relevant for applications related to model averaging and model selection and improves stability compared to the ones estimated using scoring rules in a small sample setting. Results from a simulation study reveal that the proposed Bayesian opinion pool methodology improves prediction accuracy. An application involving the Survey of Professional Forecasters demonstrates that the Bayesian opinion pool's inflation forecast competes well with the equal-weight aggregated inflation forecast published by the Federal Bank of Philadelphia. The application showcases the usefulness of the Bayesian solution in situations where traditional opinion pools fail.

Chapter two introduces a non-parametric vector autoregressive model with dynamic factor (DF-NPVAR) through a hierarchical Bayesian approach. The chapter considers the specification, identification and estimation of the DF-NPVAR model, allowing it to be efficiently

fit via MCMC algorithms. The model aims at effectively capturing dynamic relationships among variables and enabling the incorporation of extensive information sets. Issues related to model comparison and extensions to settings with autocorrelated errors and qualitative variables are also considered. In an application employing post-war US data, the DF-NPVAR model successfully identifies non-linear associations between macroeconomic variables and the dynamic factor captures the business cycle component, which aligns with officially declared recession periods.

Chapter three discusses a Bayesian estimation for the FAVAR models using the precision-based algorithm. The model is fully identified under the identification restrictions of Bai et al. (2016). The approach increases the efficiency of the Gibbs sampler and avoids slow convergence and poor mixing (Chan and Jeliazkov (2009)). This article then contrasts the Bayesian approach with the one-step and two-step estimation techniques proposed in Bernanke et al. (2005). The simulation study finds that the Bayesian approach recovers the unobservable factor in a simple FAVAR framework compared to estimation techniques proposed in Bernanke et al. (2005).

Chapter 1

Combining Density Forecasts using Bayesian Opinion Pools

1.1 Introduction and Motivation

A forecaster's outlook towards any predictive exercise is reflected in how they formulate, specify and estimate their model. The dynamics of the model depend on how the forecaster perceives and incorporates uncertainty (Steel (2020)). As a result, several competing predictive models emerge for a given random variable. For a researcher, an intuitive way to utilize all this information is by aggregating all the predictive densities (See Hoeting et al. (1999) for Bayesian Model averaging, Wang et al. (2009) for frequentist model averaging, Moral-Benito (2015) for model averaging in economics, Gneiting and Ranjan (2013) for predictive model aggregation and Clyde and George (2004) for model uncertainty). This paper focuses on linear opinion pool (Stone (1961), Bacharach (1974)), a simple and widely used method for model aggregation, and explores its utility for time series forecasting applications under the Bayesian foundation.

To set up the framework, let y_t be a random variable and $Y_T = \{y_1, y_2, \ldots, y_T\}$ be a sequence of ordered random variables up to time T. Let M_k be the model estimated by forecaster k and $p(y_{T+1}|Y_T, M_k)$ be the predictive density of y_{T+1} associated with M_k, where $k = 1, 2, \ldots, K$. The aggregated predictive density, $p(y_{T+1}|Y_T)$, under the linear opinion pool framework is obtained as

$$p(y_{T+1}|Y_T) = \sum_{k=1}^{K} w_{k,T} p(y_{T+1}|Y_T, M_k), \tag{1.1}$$

where $w_{k,T}$ is the weight allotted to M_k with the time subscript implying the use of information up to T. It also means that the weights are updated recursively once y_{T+1} is realized. The weights are estimated with respect to the constraints $\sum_{k=1}^{K} w_k = 1$ and $w_k \geq 0 \ \forall \ k = 1, 2, \ldots, K$, ensuring that Eq. 1.1 is an appropriate probability density.

Researchers have estimated Eq. 1.1 by optimising different objective functions (further referred to as traditional opinion pools or TOP). One issue with the approach is that estimating unique weights is challenging if the number of predictive models exceeds the number of observations (micronumerosity). A non-negative degree of freedom is a necessary condition for optimization-based problems (like least squares or optimization based on scoring rules). Often, micronumerosity (or near micronumerosity) becomes a binding constraint, especially in time series forecasting, where the frequency of observations limits the data length. Researchers have used different types of regularization (like ridge or lasso) which allows the estimation by shrinking some of the weights towards 0. This regularization provides one-way shrinkage towards 0 and thus forces the researcher to consider fewer models in the final analysis. Second, the simulation study found that the weights under TOP have a high variance when the sample size is small. Since weights are estimated recursively given past accuracy of forecasters, instability over time means that the data does not have sufficient information regarding the consistency of predictive performance. In these cases, weights

associated with the expert's prediction react to their immediate past forecasting accuracy which does not ensure the sustenance of accuracy in the future (the problem of overfitting). This could be one of the reasons why equal weights perform competitively with optimized weights (Hendry and Clements (2004) and Wallis (2005)), as equal weights provide insurance against bad forecasts.

The paper proposes to estimate the opinion pools using the proper Bayesian formulation and hence calls it the Bayesian Opinion Pool (BOP). This approach resolves the two issues discussed. First, the Bayesian framework allows the opinion pool to be estimated when the number of forecasting densities exceeds the number of observations. The introduction of prior compensates for the lack of observations in the data. The proposed algorithm is effective even when dealing with a high number of forecasters since the whole vector of weights is sampled in a single block, leading to computational efficiency. Second, the BOP's weights are stable under the small sample setting. The BOP utilizes the Dirichlet prior which allows the opinion pool's weight to shrink towards any possible combination rather than restricting to boundary cases like ridge or lasso. The researcher can choose to shrink the weights on the spectrum with one extreme of allotting equal weights to all the models to another where all the weights are allotted to the best model. This makes BOP useful for applications related to model averaging and model selection. Due to the flexibility, the prior can introduce stability in the weights under the small sample settings, by shrinking them towards equal weights and allowing deviations only if enough evidence is available in the data. The stability in the weights over time leads to improvement in prediction accuracy since the shrinkage avoids overfitting. The simulation study (Section 1.4) found evidence that the BOP is stable under a small sample setting and is highly competitive with the TOP (the five proper scoring rules considered are log, quadratic, spherical, continuous ranked probability score (CRPS), and the first two moments score (FTMS)). Finally, the paper uses BOP in an application involving the survey of professional forecasters (SPF) where traditional optimization-based opinion pools fail due to micronumerosity. The aggregated predictive density for the inflation rate

is estimated and compared with the equal weights strategy (simple opinion pools or SOP) published by the Federal Bank of Philadelphia. The paper finds evidence for lower MSPE associated with inflation estimated using the BOP.

Even though the current framework uses Bayesian formulation, that does not restrict the experts to Bayesian models. In the Bayesian setting, the predictive density $p(y_{T+1}|Y_T, M_k)$ can be written as

$$p(y_{T+1}|Y_T, M_k) = \int p(y_{T+1}|\theta_k, Y_T, M_k)p(\theta_k|Y_T, M_k)d\theta_k, \tag{1.2}$$

where θ_k be the set of parameters used to specify M_k, $p(\theta_k|Y_T, M_k)$ is the posterior density of θ_k and $p(y_{T+1}|\theta_k, Y_T, M_k)$ is the likelihood function associated with M_k evaluated at the value y_{T+1}. The parameter θ_k has been integrated out, so it does not appear in the Eq. 1.2. In the likelihood-based perspective, the predictive density can take the form

$$p(y_{T+1}|Y_T, M_k) = p(y_{T+1}|\theta_k^{MLE}, Y_T, M_k). \tag{1.3}$$

The density is conditioned on $\theta_k = \theta_k^{MLE}$, the maximum likelihood estimator, on the right side of Eq. 1.3, and thus, θ_k got absorbed into M_k on the left side of Eq. 1.3.

Extensive research has been done involving aggregation of the predictive densities. Mitchell and Hall (2005) combined density forecasts using Kullback–Leibler information criterion. Billio et al. (2013) used state space modelling to aggregate predictive densities and used Bayesian formulation to estimate time-varying weights. Busetti (2017) discussed quantile aggregation of predictive densities. Bassetti et al. (2018) used the Bayesian method to estimate the beta transformation of the opinion pool. McAlinn and West (2019) develop a

novel class of dynamic latent factor models for time series forecast synthesis called Bayesian predictive synthesis which encompasses several existing forecast pooling methods.

For forecasting applications, an appropriate class of objective functions used to estimate opinion pools are scoring rules. As the name suggests, a scoring rule is a function that assigns a score to a probabilistic density based on how well it predicts the realized event. Scoring rules can be judged based on *ex-ante* and *ex-post* properties (Winkler et al. (1996)). A scoring rule is called proper (an *ex-ante* property) when it disincentivizes the forecaster from revealing the probability density different from their true belief. *Ex-post* properties are concerned with how the scoring rule evaluates the performance of a probabilistic density. Gneiting and Raftery (2007) covered a thorough discussion on proper scoring rules and their theoretical properties. Bates and Granger (1969) optimized weights in Eq. 1.1 by minimizing the variance. Degroot and Mortera (1991) estimated optimal weights by minimizing the expected quadratic score under the Bayesian framework. Geweke and Amisano (2011, 2012) optimized weights using the log score and showed its usefulness via stock index data. Opschoor et al. (2017) compared opinion pools optimized from censored likelihood score (CLS), CRPS, and log score on stock market indices data and found that CLS performed the best, whereas the log score performed the worst. Garratt et al. (2023) transform the linear opinion pool using the empirical cumulative distribution function to improve the matching with the marginal density of the target variable. The properties of opinion pools vary based on the scoring rule used to estimate them.

This paper contributes to the renewed interest in survey-based measures of inflation forecasts. The Federal Reserve Bank of Philadelphia collects inflation predictive densities from several forecasters and publishes them on its website. They weigh all the predictive densities equally to calculate the aggregate level density, thus obtaining the simple opinion pool (SOP). The issue with the equal weights is that they do not utilize the past predictive accuracy of the forecasters and thus, lead to an inefficient estimator of predictive density.

This paper applies BOP to estimate weights for aggregated inflation density. The issue of ignoring past predictive accuracy information could have been tackled through TOP, but due to low data frequency, it is infeasible to use optimized-based methods for any length of the training window. The inflation forecast obtained through the BOP at various levels of shrinkage competes well with the Federal Reserve Bank of Philadelphia's published SOP. The average density allotted to realized inflation is higher, and the MSPE associated with the estimated expected value is lower for the BOP than the SOP. The paper concludes by providing future avenues of research.

Section 1.2 gives a brief overview of traditional opinion pools. Section 1.3 introduces the Bayesian opinion pool. Section 1.4 presents the simulation study where the performance of TOP and BOP are investigated in several settings. Section 1.5 covers the macroeconomic application involving the SPF data. Section 1.6 concludes the paper.

1.2 Scoring Rules for Traditional Opinion Pools

This section summarizes the asymptotic properties of TOP and how it is used to estimate opinion pools. Let θ_0 be the true vector of parameters, M_0 be the true model or DGP and $p(y_t, \theta)$ be the parametric probability distribution. Let any proper scoring rule be presented as $S(\cdot)$. Gneiting and Raftery (2007) showed that asymptotically

$$arg \ \max_{\theta} \frac{1}{T} \sum_{t=1}^{T} S(p(y_t, \theta)) \longrightarrow \theta_0 \qquad as \qquad T \longrightarrow \infty. \tag{1.1}$$

If the constraints on weights in Eq. 1.1 are satisfied the opinion pool satisfies the conditions of an appropriate probability distribution. Thus, they can be represented as a probability distribution, $p(y_t, w_T)$, where $w_T = \{w_{1T}, w_{2T}, \ldots, w_{KT}\}$ be the parameter of interest. Then

asymptotically,

$$arg \ \max_{w_T} \frac{1}{T} \sum_{t=1}^{T} S(p(y_t, w_T)) \longrightarrow w_0 \qquad as \qquad T \longrightarrow \infty. \tag{1.2}$$

Bernardo and Smith (2000) considered three possible scenarios in the context of model averaging. First is the M-closed case where M_0 is identified and available in the model list. In this case, opinion pools will converge to M_0 asymptotically and $w_0 = \{1, 0, \ldots, 0\}'$ where the weightage of 1 is allotted to M_0 and 0 to other models. The second case is when M_0 is available, but the researcher decides to intentionally leave it out of the model set (M-complete case). The third one is the M-open case, the most applicable and is considered in this paper, is when M_0 is not part of the model list. In this case, w_T will converge to some weight vector $w_0 = w^*$, which is related to the properties of the metric implied by the scoring function. For example, the log score minimizes the Kullback–Leibler divergence from M_0 to the opinion pool (Gneiting and Raftery (2007)).

Elliott et al. (2016) argued that there is no natural choice for choosing the scoring rule under the M-open case. Thus, the paper considers log (L), quadratic (Q), spherical (S), CRPS (C), and FTMS (F) scoring rules for estimating weights in Section 1.4. For a given predictive density $p(y_{T+1}|Y_T, M_k)$ and realization of $y_{T+1} = y^r_{T+1}$, these scoring rules will allot a score

as

$$
\begin{aligned}
L(y^*) &= log(p(y^r_{T+1}|Y_T, M_k)) \\
Q(y^*) &= 2p(y^r_{T+1}|Y_T, M_k) - \int_{-\infty}^{\infty} p(y_{T+1}|Y_T, M_k)^2 dy_{T+1} \\
S(y^*) &= \frac{p(y^r_{T+1}|Y_T, M_k)}{(\int_{-\infty}^{\infty} p(y_{T+1}|Y_T, M_k)^2 dy_{T+1})^{0.5}} \\
C(y^*) &= -\int_{-\infty}^{y^r_{T+1}} F(y_{T+1}|Y_T, M_k)^2 dy_{T+1} - \int_{y^r_{T+1}}^{\infty} (F(y_{T+1}|Y_T, M_k) - 1)^2 dy_{T+1} \\
F(y^*) &= -\Big(\frac{y^r_{T+1} - \mu_k}{\sigma_k}\Big)^2 + log(\sigma_k^2),
\end{aligned}
\tag{1.3}
$$

where μ_k is the mean, σ_k is the standard deviation and $F(y_{T+1}|Y_T, M_k)$ is the cumulative predictive density for M_k. The scoring rules are multiplied by minus one to have higher scores implying improved forecast performance. An issue with the CRPS rule is that the optimization becomes computationally heavier as the number of predictive densities increases due to the presence of integration (Gneiting and Raftery (2007)). Dawid and Sebastiani (1999) suggested four proper scoring rules based on the first two moments of the predictive distribution, and $F(y^*)$ is chosen to be the most popular one in this paper. Given Y_T is realized, the weights for the opinion pools are estimated as

$$
w^*_T = arg \max_{w_T} \sum_{t=1}^{T} S\Big(\sum_{k=1}^{K} w_{k,T}\, p(y_t|Y_{t-1}, M_k)\Big), \tag{1.4}
$$

where $w^*_T = \{w^*_{1T}, w^*_{2T}, \ldots, w^*_{KT}\}$. The weight vector has a T subscript instead of t to represent the information incorporation up to T and the estimation being recursive. The

opinion pool for y_{T+1} will take the form

$$p(y_{T+1}|Y_T) = \sum_{k=1}^{K} w^*_{k,T}\, p(y_{T+1}|Y_T, M_k). \tag{1.5}$$

1.3 Bayesian Opinion Pool

This section lays out the estimation procedure for opinion pools under the proper Bayesian formulation. To obtain the posterior density of weights given data, the Bayes theorem is utilized and is given as

$$\begin{aligned} p(w_T|Y_T) &\propto p(Y_T|w_T)p(w_T) \\ &\propto \prod_{t=1}^{T}\Big(\sum_{k=1}^{K} w_{k,T}\, p(y_t|Y_{t-1}, M_k)\Big)\prod_{k=1}^{K} w_{k,T}^{\alpha_k-1}. \end{aligned} \tag{1.1}$$

The weights are treated as a K-dimensional, simplex bound, random variable endowed with a Dirichlet prior.

$$p(w_t) \sim Dir(\alpha_1, \alpha_2, \ldots, \alpha_K), \tag{1.2}$$

where the hyperparameter α_k determines the relative weight given to $M_k \;\forall\; k = 1, 2, \ldots, K$, thus allowing the incorporation of prior information for any forecaster. One of the important properties of the Dirichlet Prior is its adaptability to applications related to model selection and model averaging. If the value of α_k is kept above 1 for all k, the prior penalizes allotting

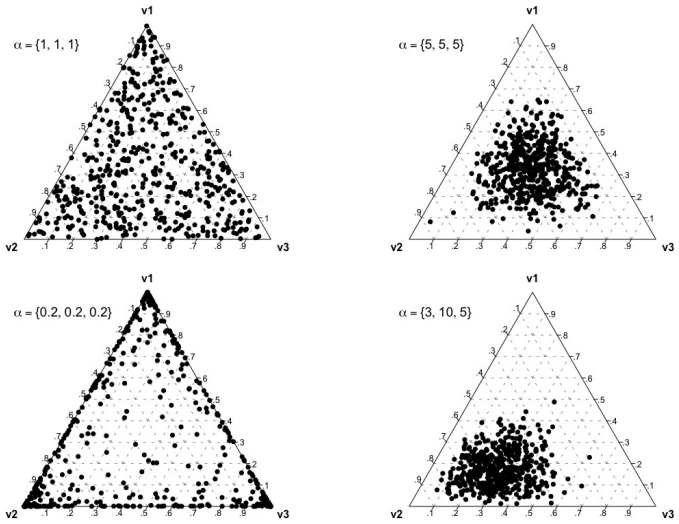


Figure 1.1: Draws from 3-dimensional Dirichlet with different α

extreme weights to some models; thus, the posterior mean of weights tends to be closer to equal weights. As α_k tends towards infinity for all k, the posterior mean of weights tends towards equality. If the value of α_k is kept below 1 for all k, the prior incentivizes extreme weights for some models. As α_k tends towards 0, the posterior weights tend towards choosing the best model among the set. This is useful in case the application requires model selection. Keeping $\alpha_k = 1$ for all k makes the prior uniform which lets the data steer the posterior mean of weights towards optimized values while imposing mild shrinkage since the prior's mean is equal weights. To illustrate, Figure 1.1 shows 4 different cases of draws from 3-dimensional Dirichlet distribution for different values of $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$. The top-left figure represents the uniform prior with $\alpha = \{1, 1, 1\}$ where all possible combinations of weights are equally likely. The top-right figure represents a stronger shrinkage of weights towards 1/3 since $\alpha = \{5, 5, 5\}$ which can be useful for model averaging. The bottom-left figure represents the shrinkage of weights towards boundary cases since $\alpha = \{0.2, 0.2, 0.2\}$ which can be useful for model selection. The bottom right figure represents the case when non-sample information is available about the experts and all experts are not preferred equally.

Given that the opinion pool itself is an appropriate probability distribution function, it makes

sense to treat it as the joint conditional density (equivalent to the joint likelihood function) given as

$$\begin{aligned} p(Y_T|w_T) &= \prod_{t=1}^{T} p(y_t|Y_{t-1}) \\ &= \prod_{t=1}^{T} \Big(\sum_{k=1}^{K} w_{k,T}\, p(y_t|Y_{t-1}, M_k) \Big). \end{aligned} \tag{1.3}$$

The conditional density incorporates the past predictive performance of experts as it is defined as a sequence of one-step-ahead conditional densities from time 1 to T. Since each conditional density is a mixture generated by the weights which are not varying with respect to time, the weights are tied with past conditional densities.

Given the prior and the conditional densities, the posterior density of the weights takes the form

$$\begin{aligned} p(w_T|Y_T) &\propto p(Y_T|w_T)p(w_T) \\ &\propto \prod_{t=1}^{T} \Big(\sum_{k=1}^{K} w_{k,T}\, p(y_t|Y_{t-1}, M_k) \Big) \prod_{k=1}^{K} w_{k,T}^{\alpha_k - 1}. \end{aligned} \tag{1.4}$$

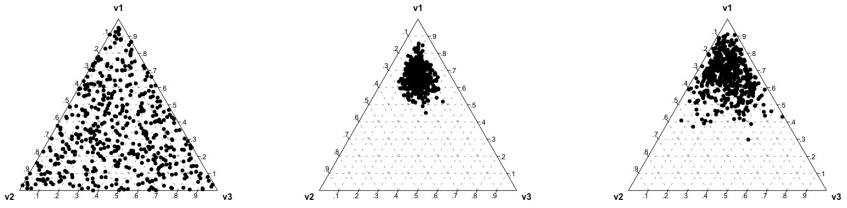
It is easy to see that the log score rule (optimal prediction pools by Geweke and Amisano (2011)) is a monotonic transformation of the conditional density.

$$log\Big(\prod_{t=1}^{T} \big(\sum_{k=1}^{K} w_{k,T}\, p(\pi_t|\pi_{1:t-1}) \big) \Big) = \sum_{t=1}^{T} log\Big(\sum_{k=1}^{K} w_{k,T}\, p(\pi_t|\pi_{1:t-1}) \Big).$$

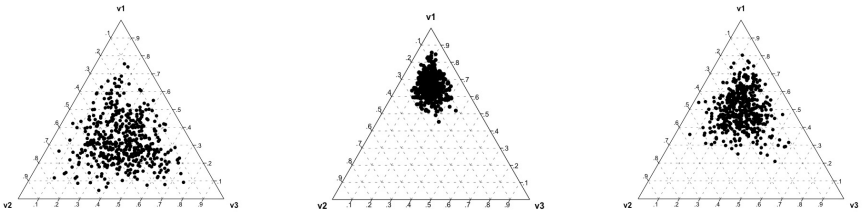
Therefore, the mean of the posterior density of weights will coincide with the weights under

the optimal prediction pool asymptotically. Gneiting and Raftery (2007) showed that the log score minimizes the Kullback–Leibler divergence from DGP to the prediction model. This means the weights under the BOP minimize the Kullback–Leibler divergence from DGP to the opinion pool since the prior disappears in a large sample (Proof in the Appendix). In small sample settings, the estimates of BOP will differ from the optimal prediction pool as the BOP weights will shrink towards the prior. Since it is not feasible to optimize the function under micronumerosity, the BOP with a uniform prior can be seen as an extension of the optimal prediction pool, broadening its applicability.

To illustrate how the prior density interacts with the conditional density, Figure 1.4 showcases how uniform prior (top figure), tight prior with $\alpha > 1$ (middle figure) and boundary prior with $\alpha < 1$ (bottom figure) shrink weights. The uniform prior expands the set of feasible vectors of weights and shifts the mean of conditional density towards equal weights. Increasing the value of α increases the strength of shrinkage (as seen in the middle figure) and thus is useful in applications requiring all models to participate in the final analysis (model averaging). The boundary prior shrinks weights towards the best model which can be useful for applications related to model selection.



Prior ($\alpha_k = 1$) + Conditional Density = Posterior Density



Prior ($\alpha_k > 1$) + Conditional Density = Posterior Density

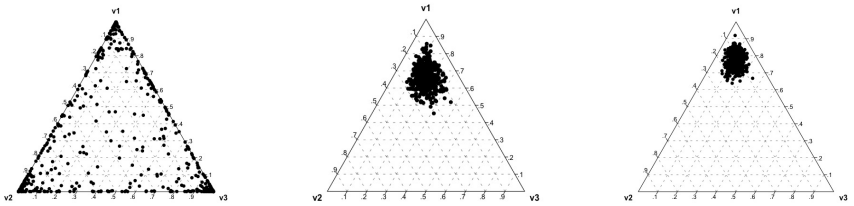


Figure 1.4: Illustration of weight's shrinkage under Bayesian opinion pool.

Since the final form of the posterior is non-standard, the paper uses the Metropolis-Hasting (MH) algorithm to draw from the posterior density. For the proposal density, one potential candidate is the Dirichlet distribution centred at the previous draw. Let $w_T^{(g)}$ be w_T drawn in the g^{th} iteration. The Markov Chain Monte Carlo (MCMC) estimation of the BOP for the Dirichlet proposal is summarized in the following steps.

STEP 1. Choose a value of $w_T = w_T^{(0)}$.

STEP 2. At the g^{th} iteration, sample $w_T^{(g)} \sim Dir(\alpha^{(g-1)})$ where $\alpha^{(g-1)}$ is chosen to center the distribution at $w_T^{(g-1)}$.

STEP 3. Generate $u \sim U(0,1)$.

STEP 4. If $u \leq min\Big(\frac{p(w_T^{(g)}|Y_T)Dir(w_T^{(g-1)}|w_T^{(g)})}{p(w_T^{(g-1)}|Y_T)Dir(w_T^{(g)}|w_T^{(g-1)})}, 1\Big)$, return $w_T^{(g)}$, else return $w_T^{(g-1)}$. Go to step 2 and continue until the desired number of iterations is obtained.

Since there is no obvious choice for $\alpha^{(g-1)}$ in step 2, the researcher can choose $\alpha^{(g-1)} = cw_T^{(g-1)}$, where c is chosen based in the rejection rate. For the Dirichlet proposal, the acceptance rate may be too low if the posterior density is narrow or the dimension is high. Alternatively, the vector of weights can be transformed to be defined on an unbounded domain using a multivariate logit transformation. Given $\theta_T = \{\theta_{1,T}, \ldots, \theta_{K-1,T}\}$, the transformation will

look like

$$\theta_{k,T} = \ln(\frac{w_{k,T}}{w_{K,T}}) \tag{1.5}$$

for all $k = 1, \ldots, K-1$. The draws are sampled from a tailored proposal normal density as $\theta_T \sim N(\bar{\theta}_T, \bar{\Omega}_T)$. The mean of the Gaussian proposal, $\bar{\theta}_T$ is kept at an optimized value calculated using a back-fitting MCMC algorithm (details can be found in the Appendix) since the numerical optimization is infeasible under micronumerosity. The covariance matrix, $\bar{\Omega}_T$ can be kept equal to either σI_{K-1} where σ is decided based on the rejection rate or proportional to the inverse Hessian of the conditional density at $\bar{\theta}_T$. Let $\theta_T^{(g)}$ be θ_T drawn in the g^{th} iteration. The MCMC estimation of the BOP for the transformed proposal is summarized in the following steps.

STEP 1. Choose a value of $\theta_T = \theta_T^{(0)}$

STEP 2. At the g^{th} iteration, sample $\theta_T^{(g)} \sim N(\bar{\theta}_T, \bar{\Omega}_T)$.

STEP 3. Transform $\theta_T^{(g)}$ to obtain $w_T^{(g)}$.

STEP 4. Generate $u \sim U(0,1)$.

STEP 5. If $u \leq min\Big(\frac{p(w_T^{(g)}|Y_T)q(w_T^{(g-1)})}{p(w_T^{(g-1)}|Y_T)q(w_T^{(g)})}, 1\Big)$, return $w_T^{(g)}$, else return $w_T^{(g-1)}$. Go to step 1 and continue until the desired number of iterations is obtained.

The density $q(\cdot)$ is the transformed density for w_T obtained after incorporating the Jacobian of the transformation. The mean of the Gaussian proposal, $\bar{\theta}_T$, can also be kept as the previous draw (just like in the case of the Dirichlet proposal).

The framework is not restrictive to one-step-ahead densities and can be extended for long horizons. For a given predictive density $p(y_{t+h}|Y_t, M_k)$ representing h step ahead forecast,

the posterior density of the weights will look like

$$p(w_T^h|Y_T) \propto \prod_{t=1}^{T} \Big(\sum_{k=1}^{K} w_{kT}^h \, p(y_{t+h}|Y_t, M_k) \Big) \prod_{k=1}^{K} (w_{kT}^h)^{\alpha_k - 1}, \tag{1.6}$$

where $w_T^h = \{w_{1T}^h, \ldots, w_{KT}^h\}$ represents weights optimized using information upto time T and used for predictions in period $T + h$.

1.4 Simulation Study

This section explores the predictive performance of BOP and TOP on simulated data. The DGP and individual models are considered under the linear setting to preserve useful insights that might get lost in a complicated analysis. Let the variable of interest be z_t. The data is artificially generated as

$$DGP : z_t = 0.5 + 0.5 z_{t-1} + \epsilon_t \,, \;\; \text{where } \epsilon_t \overset{\text{iid}}{\sim} N(0, 5). \tag{1.1}$$

Let z_t^r be the realized value of z_t at time t. There are three experts who submit their forecasts for z_t^r as $N(z_t^r, \text{var}_k)$ for $k = 1, 2$ and 3. Therefore, the mean of the expert's predictive densities are unbiased and only the variance differ as follows

	Case 1	Case 2
Expert 1	$\text{var}_1 = 4$	$\text{var}_1 = \text{AR}(1)$
Expert 2	$\text{var}_2 = 8$	$\text{var}_2 = \text{AR}(1)$
Expert 3	$\text{var}_3 = 16$	$\text{var}_3 = \text{AR}(1)$

Table 1.1: Variance for Expert's Predictive Densities

Case 1 tests the situation where there exists a clear ranking in accuracy within the experts. Expert 1 is the most accurate whereas expert 3 is the least. This scenario tests the ability of BOP and TOP to identify the best model for different training samples. Case 2 models the variance of the mean of the predictive density to follow an autoregressive process with varying degrees of persistence. This scenario introduces persistence to the accuracy of an expert's prediction while allowing the ranking of experts to change over time. Thus, one expert can predict accurately for some periods and inaccurately for others. This tests the ability of BOP and TOP to estimate the weights which optimizes forecasting accuracy while ensuring against bad predictions due to the over-reliance on any one of the experts.

The opinion pools are trained using the following sample sizes: $T \in \{5, 10, 20, 30, 50, 100\}$ for case 1 and $T = \{30, 100, 200\}$ for case 2. Case 2 is tested with a larger T allowing for rankings to change over time. The testing sample is kept at 30 observations and predictions are made for short-term (one step ahead), medium-term (three steps ahead), and long-term (six steps ahead) forecasting horizons (h). The persistence levels of 0.2, 0.5 and 0.8 are used to test the sensitivity of results in case 2. The predictive exercise uses the rolling window approach with the window length equivalent to T. The BOP is estimated using the Dirichlet prior with varying values of α_k. A total of 10 simulation runs are considered for each case.

1.4.1 Volatility of Weights under Small Sample Setting

The paper considers case 1 to study the behaviour of weights since a clear ranking of models is defined. This exercise finds that the weights under TOP have high volatility when T is small (near micronumerosity). This can be seen in Figure 1.5 which shows the weight evolution for TOP and BOP for $T = 5, 10$ and 30 for one of the simulation runs. Each row represents different types of opinion pools whereas each column represents the sample size. For $T = 5$ and 10, the weights for TOP have a high variance, especially for log, quadratic and spherical opinion pools where the vacillations are as extreme as going from 0 and 1 in one period. This indicates that TOP is relying on the immediate predictive accuracy of the expert since there is a lack of data regarding their consistency. On the other hand, BOP is much more stable, stays close to the prior and is able to identify the best model while not overly relying on it for prediction. This is due to the Dirichlet prior with $\alpha_k = 1$ which imposes sufficient shrinkage (since it is the case of near micronumerosity) on weights. As T increases, the weights deviate from equal weights as enough evidence is present about the accuracy of the concerned model. For $T = 30$, almost all the opinion pools are able to identify the best model and allot weights according to the ranking of the models.

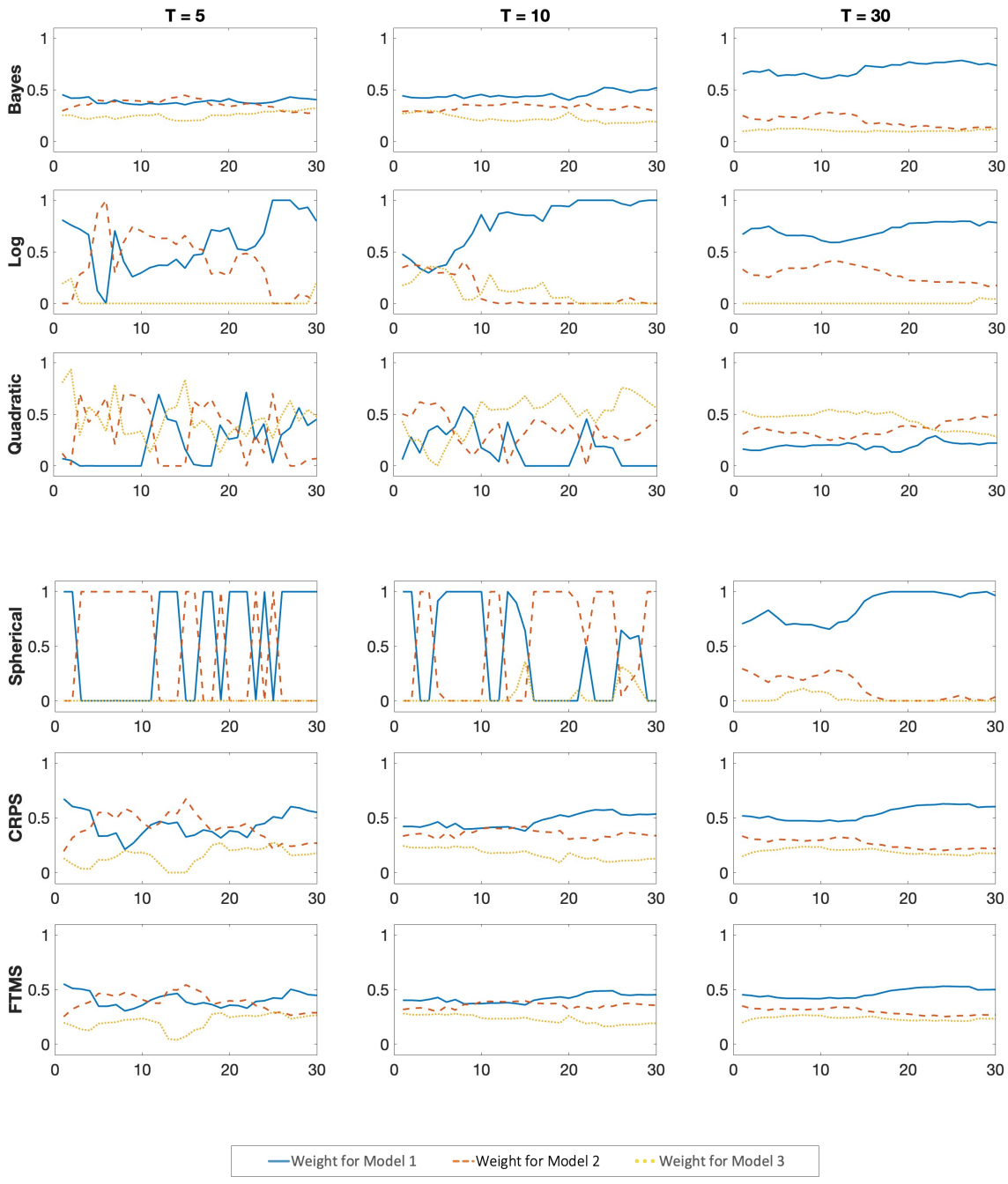


Figure 1.5: Weights evolution for opinion pools

Table 1.2 contains the summary of the standard deviation for weights for TOP and BOP. The standard deviation is estimated by first calculating the standard deviations of the weight corresponding to the individual models over the testing period and then taking the mean of

those standard deviations (three models). The BOP has the lowest standard deviation for small samples ($T = 5$ and 10) because the prior stabilizes the weights around equality. As the sample size increases ($T = 20$ and 30), the weights for TOP become stable as well and start to converge towards the best model. This is intuitive since the opinion pools are able to identify the best model when T is high.

Sample	Log	Quad	Sphere	CRPS	FTMS	Bayes
5	0.241	0.306	0.412	0.125	0.075	**0.040**
10	0.179	0.170	0.331	0.093	0.059	**0.045**
20	0.088	0.069	0.172	0.038	0.024	0.038
30	0.054	0.048	0.086	0.027	0.018	0.029

Table 1.2: Standard deviation of weights

The Dirichlet prior stabilizes the evolution of weights over time; leading to BOP having the lowest volatility under the small sample setting. The stability over time allows the opinion pool to avoid overfitting, which is one of the crucial features of a good predictive model. This positive spillover affects the predictive performance which is discussed in Subsection 1.4.3.

1.4.2 Shrinkage of Weights

Since the Dirichlet prior allows the researcher to choose the intensity of shrinkage, this subsection explores how the weight's behaviour changes as α_k changes. The paper considers case 1 since a clear ranking of models is defined. Figure 1.6 shows the evolution of weights when $\alpha_k = 1, 3$ and 5 for $T = 10, 30$ and 50. It is observed that the weights converge towards equality as α_k is increased which is an expected result. The shrinkage is strong when T is small as the prior dominates due to the insufficiency of information in the conditional density. As T increases, the conditional density starts to dominate and the weights deviate

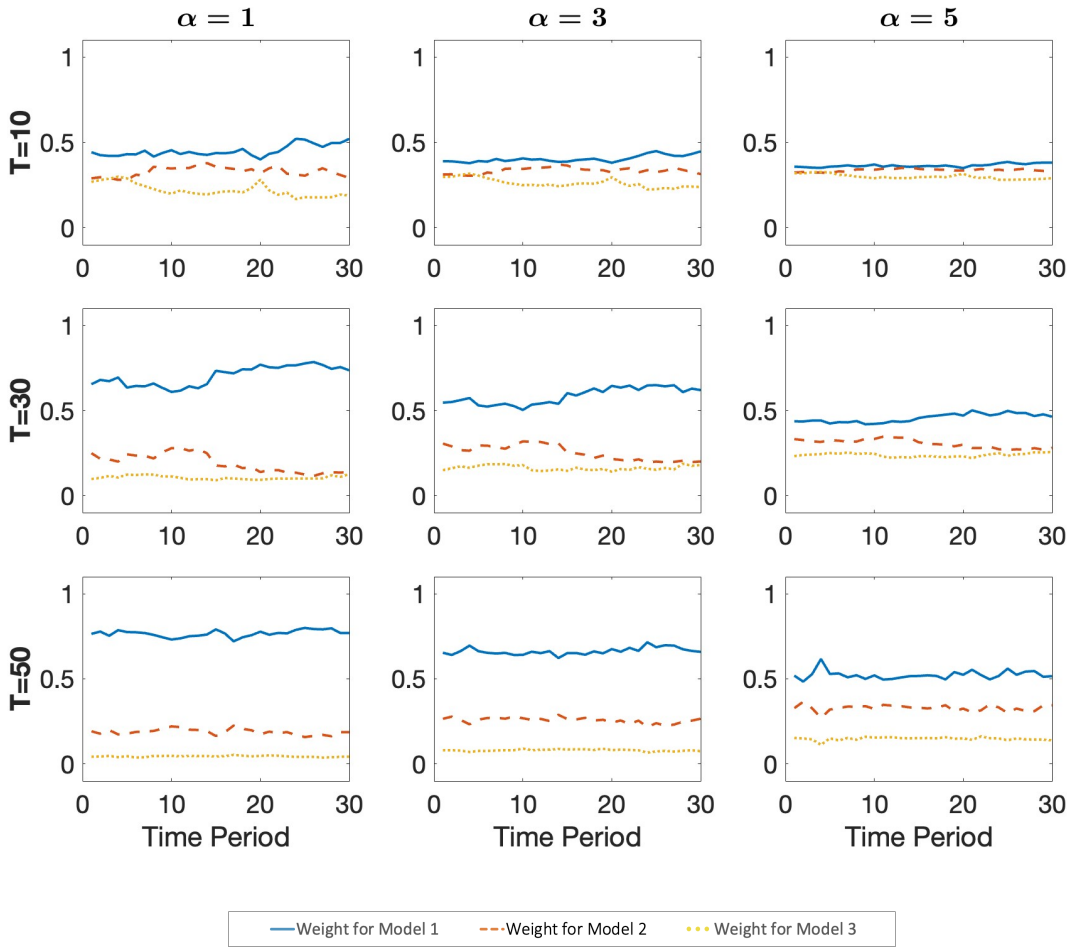


Figure 1.6: Bayesian Opinion Pool with $\alpha > 1$ for Model Averaging

from the equal weights. This property allows BOP to be used in applications related to model averaging. As α_k tends to infinity, BOP tends towards the simple opinion pool (opinion pool with equal weights).

Figure 1.7 shows the evolution of weights when $\alpha_k = 1, 0.6$ and 0.3. for $T = 10, 30$ and 50. In this case, the weights diverge away from equal weights and the best model is preferred among the available ones. As α_k tends to 0, BOP degenerates into the best model thus preferred for applications related to model selection.

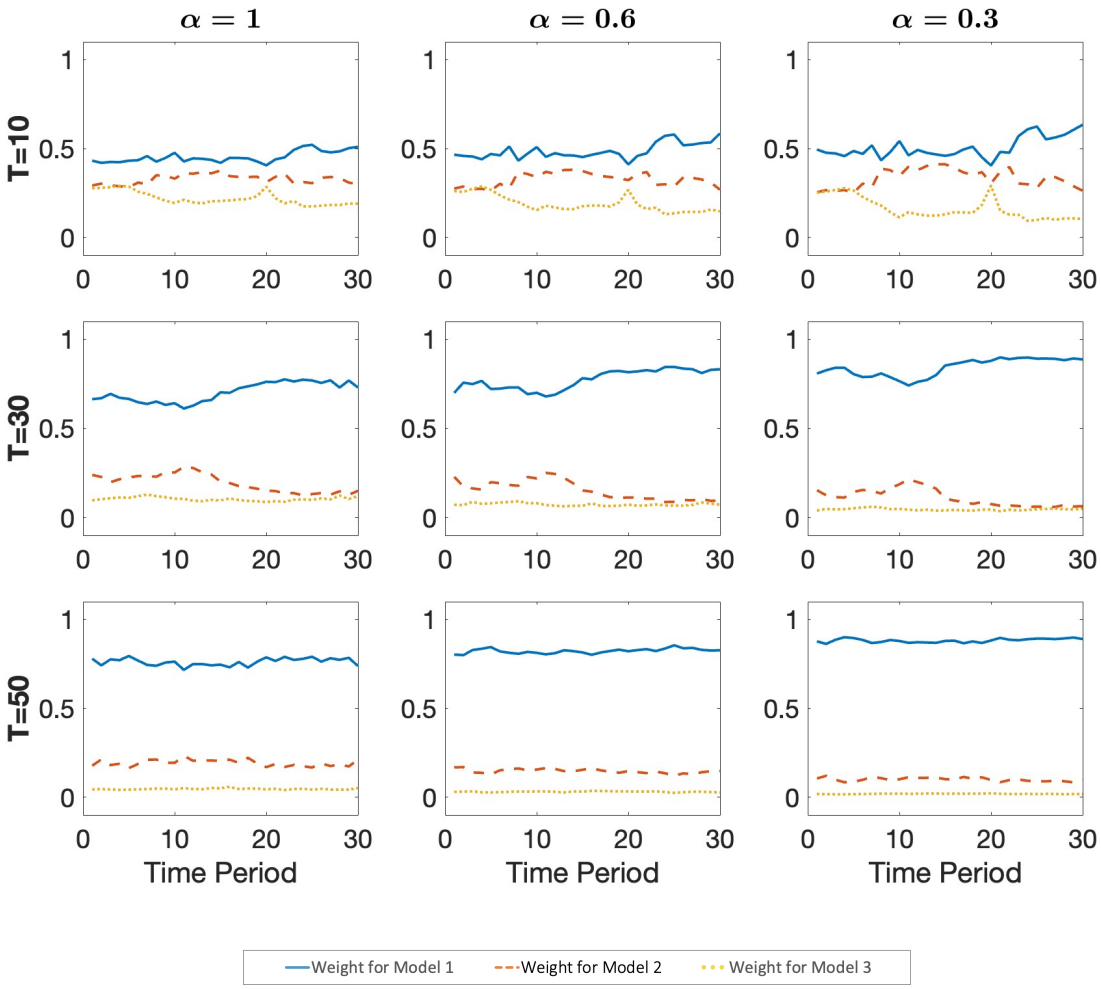


Figure 1.7: Bayesian Opinion Pool with $\alpha < 1$ for Model Selection

1.4.3 Forecasting Performance

The paper uses MSPE to evaluate the predictive performance of various opinion pools. Since decision-makers are generally interested in single-point estimates, the paper evaluates the efficacy of the opinion pools using MSPE. Table 1.3 contains MSPE values for case 1 for different sample sizes and forecasting horizons. The columns with the lowest MSPE are highlighted boldly. BOP has the lowest MSPE for most of the scenarios. One can observe that as the sample size increases, the optimal α_k also increases. Since the conditional density dominates the prior when T is high, a stronger prior leads to optimal shrinkage. Among the TOP, CRPS and FTMS perform well and their MSPE is significantly lower than log, quadratic and spherical. For FTMS, since the predictions of experts vary only through mean and variance in the DGP (normal distribution is imposed), the first two moments incorporate sufficient information for opinion pools. For CRPS, it captures the idea of proximity better than other scoring rules and thus performs well in this simulation exercise.

Table 1.4 contains MSPE values for case 2 for different sample sizes, forecasting horizons and persistence. The results are similar to that of Table 1.3 where BOP dominates for the majority of cases. CRPS and FTMS perform significantly better than log, quadratic and spherical scoring rules. This shows that BOP is able to capture the dynamic behaviour of experts when the ranking based on predictive accuracy changes over time.

1.5 Application: Inflation Prediction using the Survey of Professional Forecaster

The Survey of Professional Forecasters is a useful source of data for economists and policymakers. Croushore and Stark (2019) in "The Fifty Years of the Survey of Professional Forecasters" stated, "In 2018, the survey generated more than 45,000 unique hits to the

Sample	Horizon	Log	Quad	Sphere	CRPS	FTMS	Bayes		
							$\alpha_k = 1$	$\alpha_k = 2$	$\alpha_k = 3$
5	1	144	212	200	81.5	75.7	**74**	76.1	77.4
5	3	152	176	168	78.8	72.4	**70.5**	71.8	72.8
5	6	153	147	160	68.6	**62.5**	62.9	64.6	65.4
10	1	133	157	157	88.7	88.6	**86.9**	89.5	91.7
10	3	129	152	160	79.8	80.5	**79.1**	82.7	84.6
10	6	127	132	158	79.9	77.1	**73.5**	76.0	77.5
20	1	124	130	126	70.5	72.6	**68.1**	70.0	72.4
20	3	111	124	119	64.5	67.1	**62.5**	64.7	66.9
20	6	93.2	114	99.3	56.9	59.2	**55.2**	57.3	59.4
30	1	112	131	88.2	70.7	74.8	**67.4**	69.1	72.0
30	3	106	127	83.0	65.5	70	**62.8**	64.2	67.6
30	6	90.0	115.4	75.3	59.2	63.3	**56.3**	58.4	61.4
50	1	124	144.6	105	76.9	83.8	82.5	**76.1**	76.9
50	3	114	135	97.6	71.3	78.2	76.4	**71.0**	71.6
50	6	109	126	92.1	67.3	73.2	72.4	**67.1**	67.8
100	1	115	124	77.6	68.3	73.6	78.3	68.4	**65.4**
100	3	106	114	70.1	61	66.2	70.7	61.0	**58.1**
100	6	98.0	106	63.7	56.2	61.2	64.3	56.1	**54.1**

Table 1.3: Mean Square Prediction Error for Case 1

Persistence	Sample	Horizon	Log	Quad	Sphere	CRPS	FTMS	Bayes $\alpha_k = 1$
0.2	30	1	14.30	6.84	10.24	5.26	5.25	**5.22**
0.2	30	3	13.22	6.14	8.88	**4.60**	4.61	4.62
0.2	30	6	11.29	5.83	8.31	4.18	4.18	**4.18**
0.2	100	1	12.64	4.75	6.45	**4.12**	4.13	4.16
0.2	100	3	10.84	4.36	5.64	3.74	3.75	**3.72**
0.2	100	6	9.32	4.05	4.95	3.47	3.48	**3.45**
0.2	200	1	10.82	5.98	6.42	5.63	5.64	**5.61**
0.2	200	3	10.28	5.13	5.55	4.91	4.92	**4.91**
0.2	200	6	7.99	4.61	4.78	4.34	4.35	**4.27**
0.5	30	1	15.41	7.87	10.81	5.54	5.55	**5.51**
0.5	30	3	12.97	7.061	9.39	**4.91**	4.92	**4.91**
0.5	30	6	12.84	6.70	8.54	4.45	4.45	**4.43**
0.5	100	1	12.52	5.17	7.03	**4.45**	4.47	4.46
0.5	100	3	11.22	4.74	6.25	4.00	4.03	**3.96**
0.5	100	6	9.61	4.37	5.48	3.72	3.74	**3.68**
0.5	200	1	12.41	6.90	7.25	6.33	6.35	**6.24**
0.5	200	3	11.28	5.97	6.37	5.59	5.60	**5.53**
0.5	200	6	8.88	5.41	5.55	5.00	5.02	**4.87**
0.8	30	1	18.27	10.62	13.77	7.03	7.10	**6.98**
0.8	30	3	18.97	9.70	14.27	6.47	6.50	**6.45**
0.8	30	6	19.40	8.82	13.28	6.00	5.99	**5.94**
0.8	100	1	15.57	6.59	9.17	5.76	5.80	**5.75**
0.8	100	3	14.51	6.17	9.02	5.38	5.42	**5.34**
0.8	100	6	14.00	5.69	8.01	4.97	5.00	**4.94**
0.8	200	1	18.22	9.79	10.42	8.77	8.81	**8.61**
0.8	200	3	16.74	8.45	9.21	7.74	7.77	**7.69**
0.8	200	6	13.73	7.69	8.18	7.01	7.05	**6.86**

Table 1.4: Mean Square Prediction Error for Case 2

Philadelphia Fed's external webpages...The audience consists of academic researchers... policymakers...and business people" (P.3). Figure 1.8 shows the increase in citations and publications of papers per year which contain "Survey of Professional Forecasters" in their title, abstract or keywords[1].

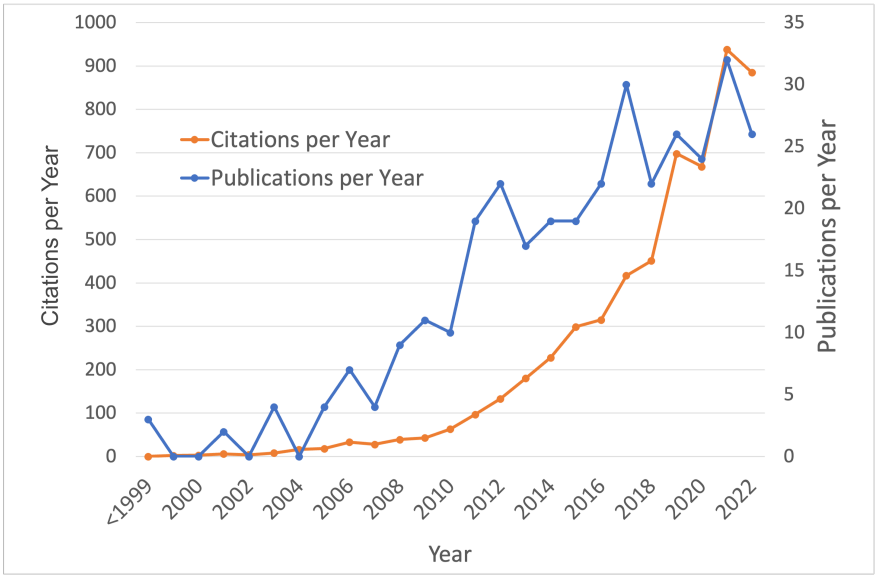


Figure 1.8: Citations and Publications per year for papers using SPF

The Federal Reserve Bank of Philadelphia publishes individual and aggregate density projections (and point estimates) for macroeconomic variables every quarter. They survey individual professional forecasters immediately after the U.S. Bureau of Economic Analysis (BEA) releases data. A unique ID is assigned to each forecaster, making tracking them possible. Anonymity is maintained to prevent strategic misreporting. The details of the data set and its significance can be found in Croushore et al. (2019), Clements et al. (2023) or on the Federal Reserve Bank of Philadelphia's website. This paper focuses on inflation density forecasts. Diebold et al. (1997) argued that point forecasts from SPF are extensively used in macroeconomic literature, but density forecasts are relatively less explored.

The experts submit their forecast densities by allotting probabilities to bins (range of inflation

[1]Survey of Professional Forecasters before 1992 was known as the ASA-NBER Survey of Expectations

rates) which are predetermined by the Fed so that the final densities are standardized and take the form of a histogram. Engelberg et al. (2009) fit continuous densities to the individual surveyed histograms to undo discretization. However, this interprets the survey replies as some subjective continuous distribution that is present in the minds of individual forecasters (Kenny et al. (2015)). Moreover, it imposes distributional assumptions which may pose practical challenges given that individual histograms are often restricted to very few bins. Thus, this paper uses linear interpolation and assumes uniform probability mass within the bins which is implicit in the assumption of a histogram (Clements (2002)).

SPF is used practically for two purposes. First, it is used to estimate inflation expectations. Inflation forecasts are integral to many macroeconomic models as they are used to estimate inflation expectations. For example, the augmented Phillips curve under aggregate price formation captures the relation where the expectations of future inflation partly drive the current inflation (Phelps (1967), Friedman (1968)). Keane and Runkle (1990) argue that a model with rational agents can be better represented using the predictive data from SPF. Coibion et al. (2018) referred to SPF extensively and argued for improved models that rely on variables with expectations. In business cycle analysis, the efficacy of a real shock depends on how much future inflation is anticipated (Kydland and Prescott (1982), Long Jr and Plosser (1983)). Under the rational expectations hypothesis, only unexpected changes in inflation lead to a change in real macro variables (Muth (1961)). The new Keynesian theory of price dynamics is based on inflation driven by its own expectations (Ball et al. (1988)). Carroll (2003) evaluated the influence of SPF data on private-sector expectations.

Second, SPF is used to forecast inflation accurately or test forecasting models. This facilitates decisions requiring accurate inflation predictions (for example, setting wage contracts). Smets et al. (2014) incorporated SPF data to measure the forecasting accuracy of New Keynesian DSGE models. Forecasts based on the neural networks and several linear econometric models were compared to SPF data (Croushore (1993)). Swanson and White (1997) used

model selection on multiple non-linear models and found that no one model was able to consistently beat SPF forecasts. Croushore et al. (2019) mentioned in their paper that "The SPF has become the gold standard for evaluating forecasts or comparing forecasting models" (P.5).

The Federal Bank of Philadelphia publishes aggregated inflation forecasts density calculated by taking a simple average of density estimates submitted by individual experts. Equal weights are a reasonable choice if the objective is to track inflation expectations. Since, the aim is to capture how rational agents perceive future inflation, including everyone's opinion captures the idea of how the economy expects inflation to be. Also, numerical optimization is infeasible as 160 forecasters participated during 120 quarters (Q1 1992 to Q4 2021), with an average of 35 active forecasters per quarter. The number of forecasters is always higher than the number of data points for any window length.

If the objective of SPF is inflation forecasting, then equal weights are a sub-optimal choice. Aastveit et al. (2018) mentioned that "Despite the long history of the SPF, little attention has historically been paid to how the weights on the competing forecast densities in the finite mixture should be determined" (P.10). The issue with the simple opinion pools (SOP) approach is that it does not exploit the information about the past predictive performance of the experts. Figure 1.9 presents the predictive performance of experts who are active for at least 10 quarters in the period of Q1 1992 to Q4 2021. The vertical axis represents the probability allotted by an expert to the bin which contained the realized value of the inflation rate. Thus, higher the probability allotted by the expert, better the forecast. The horizontal axis represents the unique ID of experts. The size of the points represents the number of quarters, an expert was active in the past. The figure depicts that some experts were consistently active and allotted much higher probability to the realized inflation rate than the average and vice versa. Using equal weights ignores this information and thus there is an opportunity to improve the predictive accuracy of aggregated inflation forecast density.

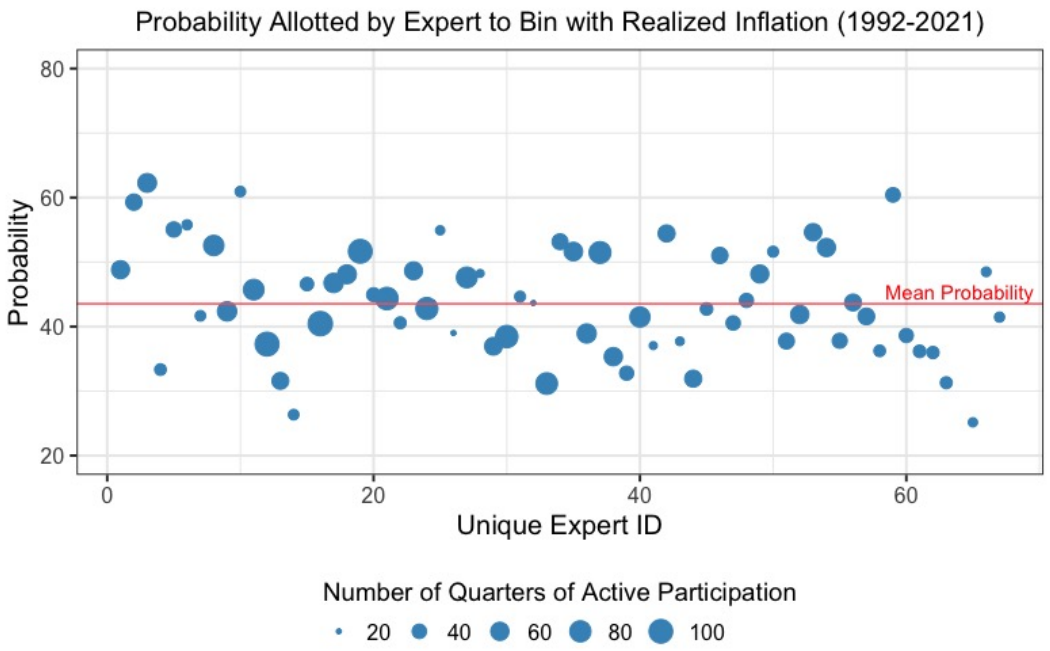


Figure 1.9: Expert's Performance in Inflation Prediction

This paper aggregates inflation density forecasts using the BOP with $\alpha_k = 1, 1.5$ and 3. BOP not only accounts for important information related to the past accuracy of individual forecasters but also allows estimation under micronumerosity where optimization-based methods fail. The decision to choose prior which shrinks weights towards equality is guided by the non-sampling information. Since the Fed uses equal weights to aggregate densities, it can be considered a good benchmark to start from. Also, researchers in the past have frequently found combining point forecasts with equal weights to be very competitive with the more complicated weighting techniques. Clemen (1989) shows in his review that equal weights are difficult to beat. Similar results were concluded by Stock and Watson (1999) and Fildes and Ord (2002). The prior shrinks the BOP towards SOP but still allows deviations in case strong evidence for better relative predictive accuracy is present.

Frequent entry and exit of forecasters make optimization of the opinion pool more involved. Capistrán and Timmermann (2009) elaborated on the problem of having an unbalanced panel

and recommended filling in the missing values before aggregation. They also considered using the unbalanced panel by keeping only the frequent forecasters. However, they had to resort to the simple average when there were fewer remaining forecasters than parameters to be estimated.

This paper does not fill in for the missing forecasting density and follows the following method to deal with the unbalanced panel.

- Entry: Suppose a forecaster is unavailable in the training data (m quarters moving window) but submits the prediction for the $(m+1)^{th}$ quarter. Thus, there is no information on the past predictive performance. In that case, their density is allotted $1/A$ weight (equal weight), where A is the number of active forecasters in the $(m+1)^{th}$ quarter. Alternatively, the researcher can choose to include the expert only if they have participated for a certain number of quarters (Conflitti et al. (2015) used 5 quarters of data).

- Exit: Suppose a forecaster was available in the training data (m-quarter moving window) but not for the $(m+1)^{th}$ quarter. In that case, their density will be allotted 0 weight, and they will not be considered in the optimization process.

- Partial Availability: Suppose a forecaster submits the prediction for $(m+1)^{th}$ quarter but was available in s periods out of m training period where $s < m$. The weights associated with the forecaster will enter the joint conditional density (Eq. 1.3) in the periods where they were available (total of s times). Thus, the methodology rewards consistency as the forecaster with active participation will have a greater influence on the opinion pool density than an inactive one.

To explain it better, let us assume that 40 forecasters were active in the last 20 quarters (not necessarily for every quarter), which is the training period for this case. Only 10 forecasters

submitted their predictions for the 21^{st} quarter, including 2 new ones. Then, the weights allotted to these 2 new ones would be 1/10 each, and the weights for the remaining 8, whose values were estimated based on the past data (excluding the twelve inactive forecasters), would be normalized so that the total sum of the weights for 10 active experts is 1.

The application considers moving windows approach with varying lengths from 21 to 29 quarters of training data and the rest of the period until 2021 Q4 as testing data. The prior with $\alpha_k > 1$ tends to bring weights closer $\frac{1}{K}$ where K is the total number of active experts.

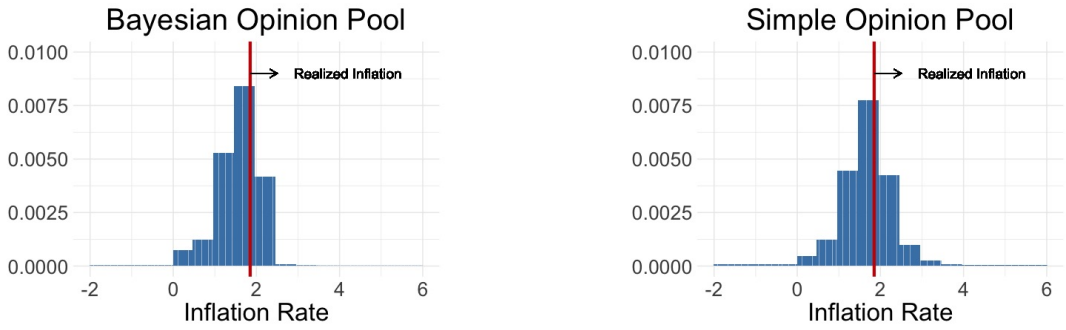


Figure 1.10: Simple and Bayesian opinion pool densities for Q1 2014

To visually aid the understanding of the information aggregation process under the BOP (uniform prior) and SOP, Figure 1.10 demonstrates BOP and SOP final predictive densities for inflation for Q1 2014. A total of 40 forecasters submitted their predictions. The final densities look different representing the fact that BOP allotted different weights to each expert than SOP. In this particular case, BOP allots a higher probability to the realized inflation than SOP

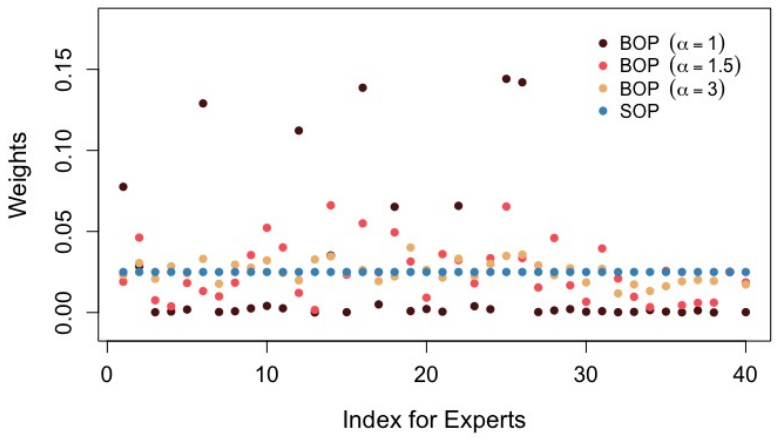


Figure 1.11: Weights allotted to experts under BOP and SOP for Q1 2014

Figure 1.11 shows estimated weights under BOP for different values of α_k for Q1 2014. The weights become concentrated around equal weights when α_k increases, implying strong shrinkage and convergence towards SOP.

The aggregated predictive densities, representing out-of-sample forecasts, are tested based on the average density allotted to the realized inflation rate. Figure 1.12 shows the average density difference between SOP and BOP evaluated at the realized inflation rate for different rolling windows. The difference is normalized to 0 and thus the vertical red line at the origin is represented by SOP. As the training window varies, the predictive accuracy of opinion pools is tested on the corresponding remaining quarters of data until Q4 2021, and the weights are updated each quarter (recursive). The average density allotted by the BOP is always higher than the SOP. As α_k increases, the MSPE difference between SOP and BOP becomes smaller representing that BOP is tending towards SOP. The difference is significant at 5% for the uniform prior for all training window lengths considered.

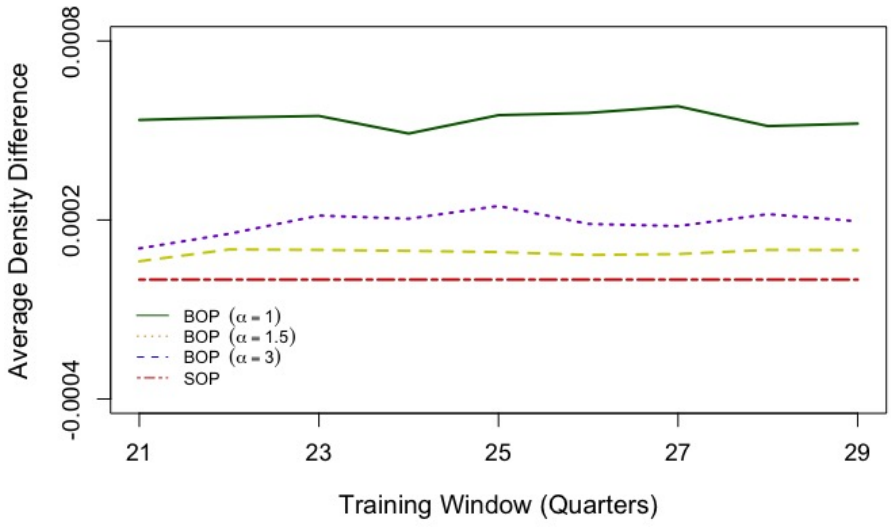


Figure 1.12: Average Density Evaluated at the Realized Inflation Rate for Opinion Pools

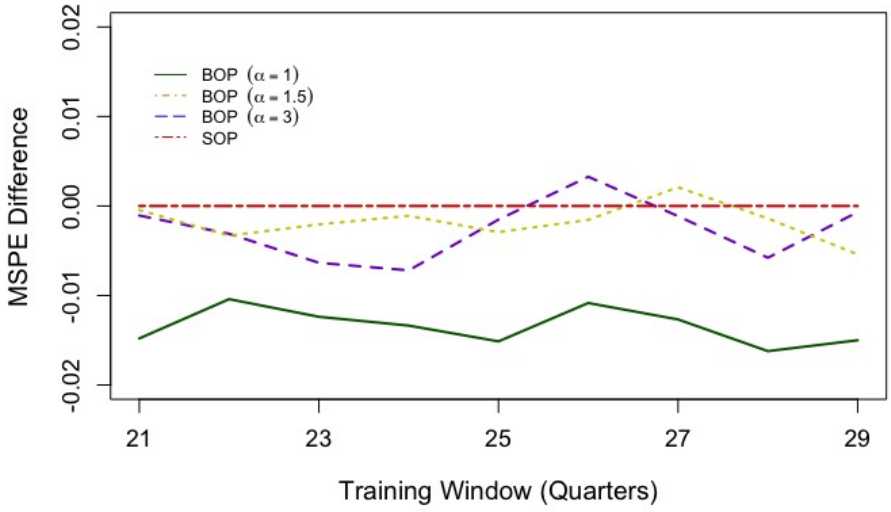


Figure 1.13: MSPE associated with opinion pools for different training windows

	21	22	23	24	25	26	27	28	29
BOP	0.433	0.440	0.445	0.453	0.451	0.457	0.462	0.467	0.468
SOP	0.449	0.454	0.459	0.463	0.467	0.471	0.476	0.482	0.488
Diff	-0.016***	-0.014**	-0.014**	-0.01*	-0.016***	-0.014***	-0.014**	-0.015***	-0.02***

Significance: 0.1 (*), 0.05 (**) and 0.01 (***)

Table 1.5: MSPE associated with opinion pools for different training windows

To test whether the improvement in the density forecast estimator spills over to the estimates of point forecasts, opinion pools are compared using MSPE. Table 1.5 presents MSPE for BOP with a uniform prior ($\alpha_k = 1$) and SOP. The MSPE for BOP is smaller than SOP for every length of training window considered. The difference in MSPE between BOP and SOP is significant. Figure 1.13 shows the MSPE difference (normalized to 0) between BOP and SOP for different training windows. The figure also shows that the MSPE difference for BOP with tight priors ($\alpha_k = 1.5$ and 3) is smaller. This validates that the complete information incorporated in estimating the forecast density has a positive spillover effect on the estimate of the point forecast. It also suggests that all forecasters are not equal in their predicting abilities, and the BOP exploits this asymmetry to its advantage.

1.6 Conclusion

This paper considers two limitations associated with traditional opinion pools. First, optimizing opinion pools using a scoring function is not feasible if the number of models exceeds the data length, an issue faced in macroeconomics applications. Second, high volatility in weight associated with TOP affects the predictive accuracy in small samples. Thus, a lot of researchers resort to equal weights since the marginal gains from optimized weights do not justify the cost of involving oneself in a complicated procedure.

This paper proposes estimating opinion pools under the Bayesian framework to resolve these issues. The Bayesian formulation allows the weights to be estimated under micronumerosity. The MCMC algorithm enables sampling of the high dimensional weight vector from its joint posterior density leading to efficiency gains. The use of a Dirichlet prior makes the weights relatively more stable over time and allows the researcher to control the shrinkage level. Apart from being feasible and flexible, the BOP is found to be highly competitive compared to TOP, when micronumerosity is not the case.

In the application, the paper uses SPF data to obtain better estimates of inflation forecasts using the BOP. The Federal Bank of Philadelphia estimates the aggregate inflation density by allotting equal weights to all the individual densities. The BOP uses optimized weights based on the past accuracy of the forecaster, thus utilizing richer information. The application showed that the BOP (with various levels of shrinkage) outperforms the SOP published by the Federal Bank of Philadelphia.

The applications of BOP extend to macroeconomics or finance, especially in settings which deal with aggregating predictive densities. Gneiting and Ranjan (2013) combined predictive cumulative distributions and tested the approach on forecasting S&P 500 returns. McAlinn et al. (2020) used the Bayesian predictive synthesis for applications related to macroeconomic forecasting. Del Negro et al. (2016) estimated time-varying weights in linear opinion pools (Dynamic Pools) and used them to investigate the relative forecasting performance of dynamic stochastic general equilibrium (DSGE) models with and without financial frictions for output growth and inflation. Baştürk et al. (2019) combined density forecasts to improve portfolio strategies.

Given that the BOP can be applied to settings involving micronumerosity, where TOP can not be estimated, and its predictive accuracy is highly competitive with TOP when micronumerosity is not an issue, a stronger case can be made for its exploration and adoption in future research. While the discussion in the paper focused on macroeconomic time series

data, the usefulness of the techniques can be extended to other applications like gambling, stock market, election polls etc. The utility of the BOP in other simulation settings, improvements in the MCMC algorithm and estimation of optimal shrinkage can be explored in future research work.

Chapter 2

Extending the Versatility of Non-Parametric VAR Models: Bayesian Specification, Estimation and Dynamic Factor

2.1 Introduction

Vector Auto-Regressive (VAR) models have significantly contributed to empirical macroeconomics literature, following the seminal work of Sims et al. (1986). The linear form is intuitive, capturing the dynamics and inter-dependency of multiple macroeconomic variables. Its popularity stems from its broad applicability, encompassing techniques such as time series modelling and forecasting. Let $y_t = \{y_{1t}, y_{2t}, \ldots, y_{Qt}\}'$ be a Q dimensional vector for all $t = 1, \ldots, T$. For a given t, the basic VAR model associates each variable with their

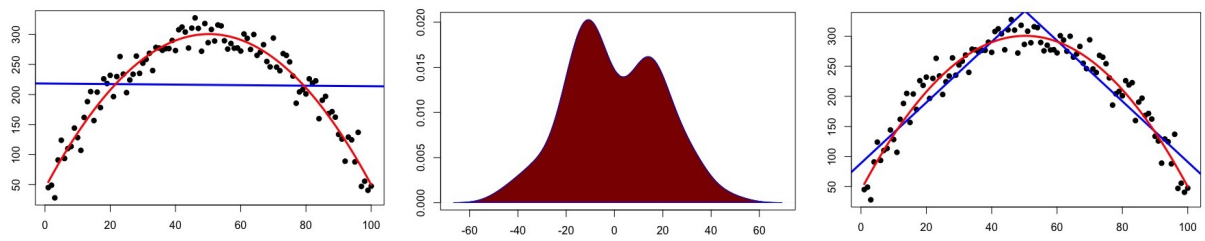
own and other variable's lag values as

$$y_t = c + \sum_{p=1}^{P} B_p y_{t-p} + \epsilon_t. \tag{2.1}$$

where c is a $Q \times 1$ dimensional vector of intercepts, B_p is $Q \times Q$ dimensional matrix of parameters for y_{t-p} and $p = 1, \ldots, P$ being the number of lags. It assumes $E(\epsilon_t) = 0$ and $var(\epsilon_t) = \Omega$ with Ω being a $Q \times Q$ dimensional variance covariance matrix.

Beyond its basic form, the VAR model can capture complicated relations between the variables. For instance, following Hamilton (1989), a lot of research has been done on regime-switching models (Hansen (1992), Chib (1996), Vigfusson (1997), Kim and Nelson (1998), Kim et al. (2005), Sims and Zha (2006)). Tong (1978) explored structural instability in univariate models and Balke (2000), Huang et al. (2005) and, Van Robays (2016) furthered it to VAR models. The time-varying parameter VAR (TVP-VAR) models relaxed the assumption of the static parameter value, allowing the coefficients to vary over time as $\{B_j t\}_{j=1}^{p}$ (Canova (1993), Cogley and Sargent (2005), Primiceri (2005), Chan and Jeliazkov (2009) and many others).

One theme common to many extensions, including those stated above, is the assumption of linearity – allowing the model to be tractable in terms of estimation and interpretation. However, the linear model is often unable to explain all the macroeconomic relations, especially when the non-linearity is prominent (Engle (1982), Robinson (1983) and Vieu (1995)). To illustrate, the paper considers the following example to motivate the types of misspecification that arise from assuming a linear form for a non-linear relation. Let the data generating process (DGP) be $y_t = 50 + 10x - 0.1x^2 + \epsilon_t, \epsilon_t \sim N(0, 20)$, which is presented in panel (A) of Figure 2.1. If linear regression is used for estimation (represented by the blue line in panel (A) of Figure 2.1), the regression residuals will be perceived as heteroskedastic which



(a) (b) (c)

Figure 2.1: Consequences of ignoring nonlinearity

was not the case in the DGP. In time series analysis, when the covariate could be a lag of the outcome variable, the misspecification can lead to the error showing serial correlation. Furthermore, even though the DGP have Gaussian errors, they will appear non-Gaussian in a linear regression (panel (B) of Figure 2.1). Due to the linearity restriction, the error distribution will be perceived as a mixture of normals and in this particular case, a bimodal. The researcher will conclude that the Gaussian assumption is inadequate when the real culprit is neglected nonlinearity.

Let's assume that the researcher chooses to approximate the nonlinearity by restricting attention to the class of piecewise linear polynomials. if a bilinear model is fit to the data (panel (C) of the Figure 2.1) one can erroneously conclude that there is evidence of structural instability. Variations away from linear models may appreciably improve the fit, although one should bear in mind that these would be spurious findings of instability or structural breaks since the underlying data-generating process is a stable, although nonlinear function. Such spurious instability, unfortunately, is not the only pitfall that can be induced by this type of misspecification.

In response to a growing need, several non-parametric/non-linear time series models were introduced. These include but are not limited to the threshold autoregressive (TAR) model by Tong and Lim (1980) and the exponential autoregressive model by Haggan and Ozaki (1981). Following in the footsteps, the assumption of non-linearity was extended to the VAR models by Härdle et al. (1998b), who estimated the non-parametric version using

polynomial fitting. Hastie and Tibshirani (1990) introduced generalized additive models, which gave structure to equations having multiple non-linear functions. Jeliazkov (2013) used this generalized additive model and estimated the non-parametric VAR (NPVAR) model under the Bayesian hierarchical framework. Kalli and Griffin (2018) estimated the non-parametric VAR model by modelling the stationary and transition densities using Bayesian non-parametric methods.

VAR models have been extensively used to characterize the joint dynamics of economic variables. The residuals of these equations are combinations of the underlying structural economic shocks, assumed to be orthogonal to each other. The structural representation requires moving from the VAR innovations to the structural shocks is called structural VAR (SVAR) analysis. It involves two steps, first being the derivation of the structural shocks from the innovations, and second that there be some economic rationale justifying it. The first step is equivalent to a no omitted variable bias: any relevant variable containing information about a structural economic shock distinct from what is already included in the VAR should be included in the analysis. Omitting that variable means that the VAR innovations will not in general span the space of the structural shocks, so the structural shocks cannot in general be deduced from the VAR innovations.

One limitation of the VAR model is its inability to incorporate many variables due to the dynamic inter-dependency structure. Every extra variable introduced in a VAR model consumes $2qp + 1$ degrees of freedom. Thus, VAR methods fail when there are hundreds of economic time series variables that potentially contain information about these underlying shocks. To deal with it, Bernanke et al. (2005) introduced the FAVAR model, which compresses the information from a large number of variables into a few factors, thus not steeply penalizing the degrees of freedom. The premise of the dynamic factor model (DFM) is that there are a small number of unobserved common dynamic factors that produce the observed comovements of economic time series. These common dynamic factors are driven by the

common structural economic shocks, which are the relevant shocks that one must identify for the purposes of conducting policy analysis. Even if the number of common shocks is small, because the dynamic factors are unobserved this model implies that the innovations from conventional VAR analysis with a small or moderate number of variables will fail to span the space of the structural shocks to the dynamic factors. Instead, these shocks are only revealed when one looks at a very large number of variables and distils from them the small number of common sources of comovement.

Bai et al. (2016) extended the FAVAR literature by studying the identification restriction and proposing a two-step sequential estimation technique. Amir-Ahmadi and Uhlig (2009) estimated the FAVAR model using the MCMC methods under the Bayesian likelihood-based approach. Another innovation explored multivariate time series models with dynamic factors (Geweke (1977), Sargent et al. (1977). Stock and Watson (2011) summarizes the three major categories of estimation approaches used in the literature for Dynamic factor models.

Though the extension of the VAR model allows for non-linearity and factors, no attempt has been made to include both in the same model. One of the reasons could be that the model estimation becomes more involved when multiple extensions are incorporated. This paper proposes a Bayesian hierarchical framework through which the VAR model with dynamic factor can be estimated under the non-parametric framework. In other words, the paper makes the first attempt to estimate the non-parametric VAR model with dynamic factors and thus calls it the DF-NPVAR model.

The remainder of the paper is structured as follows. Section 2.2 discusses the hierarchical framework of the DF-NPVAR model. Section 2.3 discusses the identification restrictions required for estimating loadings, factors and unknown additive functions. Section 2.4 presents the prior distributions used for each parameter in the model. Section 2.5 lays out the estimation procedure, which uses an efficient fitting algorithm based on MCMC simulation techniques. Section 2.6 explains the model comparison procedure. Section 2.7 suggests some

of the model extensions, including serially correlated error and the inclusion of qualitative variables. Section 2.8 considers applying the DF-NPVAR model to the post-war US economic data and making inferences from the results. Section 2.9 concludes this article.

2.2 Hierarchical Model Framework

The paper proposes a hierarchical structure for the dynamic system of Q regression equations under the Gaussian state space framework. Let the q^{th} equation is modelled through the additive form as in Hastie and Tibshirani (1990). For all $t = 1, 2, \ldots, T$, $p = 1, 2, \ldots, P$ and $q = 1, \ldots, Q$,

$$\begin{aligned} y_{1t} &= \sum_{q=1}^{Q} \sum_{p=1}^{P} g_{1qp}(y_{q,t-p}) + a_1' f_t + \epsilon_{1t} \\ \vdots \qquad & \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ y_{Qt} &= \sum_{q=1}^{Q} \sum_{p=1}^{P} g_{Qqp}(y_{q,t-p}) + a_Q' f_t + \epsilon_{Qt}, \end{aligned} \tag{2.2}$$

where y_{qt} is the dependent variable indexed by q, representing the equation it belongs to at time t and g_{rqpt} is a function in equation r evaluated at the q^{th} variable at time t with lag p, where $r = 1, \ldots, Q$. The factor f_t is a $D \times 1$ dimensional vector which enters the model only as a covariate along with a_q, which is a $D \times 1$ vector of factor loadings. The paper imposes

the Gaussian distribution on the error term where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{Qt})' \sim N(0, \Omega_1)$ with

$$\Omega_1 = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1Q} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{Q1} & \omega_{Q2} & \cdots & \omega_{QQ} \end{bmatrix}.$$

The factors follow an AR(1) process, which is described using the transition equation as follows

$$f_t \ = F f_{t-1} + v_t. \tag{2.3}$$

The error term v_t follows a Gaussian distribution given as $v_t \sim N(0, \Omega_2)$ where $\Omega_2 = diag\{\sigma_1^2, \ldots, \sigma_D^2\}$ is a $D \times D$ variance covariance matrix. The model parameter $F = diag\{\gamma_1, \ldots, \gamma_D\}$ is a $D \times D$ matrix containing coefficients for the lag factor terms.

The paper rewrites the model using Jeliazkov (2013) notations to simplify further and facilitate estimation. Let $j = qp$ be the new index for lagged variables on the right-hand side of Eq. 2.2 where $j = 1, \ldots, J$ with $J = QP$. The functions $g(.)$ are now evaluated at the new notation s_{qjt} representing $y_{q,t-p}$. The q^{th} equation for a given period t in Eq. 2.2 will look like

$$y_{qt} = g_{q1}(s_{q1t}) + \cdots + g_{qJ}(s_{qJt}) + a_q^{'} f_t + \epsilon_{qt}. \tag{2.4}$$

Equation 2.4 is stacked over time and written in the matrix form to facilitate the joint

sampling of functions and parameters from their posterior distributions. For each function in Eq. 2.4, let the T observations in the covariate vectors $s_{qj} = \{s_{qj1}, \ldots, s_{qjT}\}'$ determine the corresponding $m_j \times 1$ design point vector $v_{qj} = \{v_{qj1}, \ldots, v_{qjm_j}\}'$ with entries equal the unique ordered values of s_{qj}, that is $v_{qj1} < \ldots < v_{qjm_j}$. Let the corresponding function evaluation vector be denoted by $g_{qj} = \{g_{qj}(v_{qj1}), \ldots\ , g_{qj}(v_{qjm_j})\}'$. The matrix form for equation q in Eq. 2.2 can be written as

$$y_q = Q_{q1}g_{q1} + Q_{q2}g_{q2} + \cdots + Q_{qJ}g_{qJ} + A_q f + \epsilon_q, \tag{2.5}$$

where $A_q = I_T \otimes a_q'$ is a $T \times DT$ dimensional matrix, $\otimes$ denotes the Kronecker product, $f = \{f_1, \ldots, f_T\}'$ is a $DT \times 1$ vector, $y_q = \{y_{q1}, y_{q2}, \ldots, y_{qT}\}'$ and $e_q = \{e_{q1}, e_{q2}, \ldots, e_{qT}\}'$. The incidence matrix Q_{qj} is $T \times m_j$ dimensional which establishes the relationship between s_{qj} and v_{qj} with entries $Q_{qj}(t, m) = 1$ if $s_{qjt} = v_{qjm}$ and 0 otherwise for $m = 1, \ldots, m_j$. As there may be repeating values in s_{qj}, m_j is less than or equal to T.

The paper attempts to motivate the non-parametric functions to be perceived as realizations from a stochastic process, which tries to capture the degree of local variation between neighbouring points. A smoothness hyper-parameter is introduced in the prior of the non-linear functions, which completes the hierarchical structure of the DF-NPVAR model. The vector of function evaluations g_{qj} conditional on the smoothness parameter, τ_{qj}^2, follows a Gaussian distribution as

$$g_{qj}|\tau_{qj}^2 \sim N(g_{qj0}, \tau_{qj}^2 K_{qj}^{-1}), \tag{2.6}$$

where K_{qj} is an $m_j \times m_j$ matrix, which will be discussed later. From a Bayesian perspective, τ is interpreted as a smoothness parameter which introduces a penalty to local variation

between successive elements of g_{qj} and still allows for any possible value that the elements of g_{qj} can take. From a likelihood-based perspective, it can be interpreted as a roughness penalty term in penalized likelihood estimation (Wahba (1978)).

To complete the DF-NPVAR framework, Eq. 2.3 is stacked over time, and the complete model looks like

$$\begin{aligned} y_1 &= Q_{11}g_{11} + Q_{12}g_{12} \cdots + Q_{1J}g_{1J} + A_1 f + \epsilon_1 \\ &\vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ y_Q &= Q_{Q1}g_{Q1} + Q_{Q2}g_{Q1} \cdots + Q_{QJ}g_{QJ} + A_Q f + \epsilon_Q \\ H_f f &= v, \end{aligned} \tag{2.7}$$

where

$$H_f = \begin{bmatrix} I_D & 0 & \cdots & 0 \\ -F & I_D & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -F & I_D \end{bmatrix} \text{ and } v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}.$$

Following the Gaussian structure, the error term $v \sim N(0, S)$ where $S = diag\{\Omega_S, \Omega_2, \ldots, \Omega_2\}$. The matrix Ω_S is discussed in detail in the next section. Equation 2.7 will be used to characterize the likelihood function, which will facilitate the calculation of the conditional distribution of non-parametric functions and parameters.

2.3 Identification Restrictions

Bayesian models with proper priors do not suffer from the problem of identification (Lindley (1971); Poirier (1998)). However, due to the additive nature of non-parametric functions and the interaction of factors with loadings, there are two identification issues. The first one is associated with the intercept term as all rows of every incidence matrix sum to 1, leading to the problem of perfect collinearity among g_{qj}. In simple terms, the functions g_{qj} are correlated by construction since they enter the mean function additively which leads to the emergence of free constants. To see the problem, for any two functions r and s from Eq. 2.4, the additive form can be written as

$$\begin{aligned} g_{qr} + g_{qs} &= (g_{qr} + \alpha) + (g_{qs} - \alpha) \\ &= g^*_{qr} + g^*_{qs}, \end{aligned} \tag{2.8}$$

where $g^*_{qr} = g_{qr} + \alpha$ and $g^*_{qs} = g_{qs} - \alpha$. Since both forms are observably the same, it is obvious that neither an intercept nor the level of the individual functions is identified. There are multiple ways to anchor the functions, which will lead to their unique estimation. This paper adopts the strategy from Jeliazkov (2013) where the additive functions in q^{th} equation of Eq. 2.7 are identified by centering the $J-1$ functions as

$$y_q = Q_{q1}g_{q1} + M_0Q_{q2}g_{q2} + \cdots + M_0Q_{qJ}g_{qJ} + A_qf + \epsilon_q, \tag{2.9}$$

where M_0 is a $T \times T$ symmetric idempotent mean-differencing matrix defined as

$$M_0 \;=\; \Big(I_T - \frac{l_T l_T'}{T}\Big).$$

Centering allows the first function to capture the intercept of the regression equation.

The second identification problem is associated with the factors and loadings. Since both are unknown, the issue is related to the scale and sign identification. For any $D \times D$ matrix M_f, the problem can be stated as

$$a_q^{'} f_t = (a_q^{'} M_f)(M_f^{-1} f_t) = a_q^{*'} f_t^*, \tag{2.10}$$

where $a_q^{'*} = a_q^{'} M_f$ and $f_t^* = M_f^{-1} f_t$. Since both forms are observably identical, it is impossible to identify the scale or sign of the term uniquely. Before introducing the identification restriction, let

$$A_f = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_Q' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1D} \\ a_{21} & a_{22} & \cdots & a_{2D} \\ \vdots & \vdots & & \vdots \\ a_{Q1} & a_{Q2} & \cdots & a_{QD} \end{bmatrix}.$$

There are D^2 free terms in the M_f matrix. The model specifies Ω_2 to be a diagonal matrix (discussed earlier) which imposes $D(D-1)/2$ restrictions on Eq. 2.3. The identification problem is resolved by fixing the top D^2 part of the matrix A_f to be a lower triangular

matrix with diagonal elements restricted to 1 as

$$A_f = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{D1} & a_{D2} & \cdots & 1 \\ a_{(D+1)1} & a_{(D+1)2} & \cdots & a_{(D+1)D} \\ \vdots & \vdots & & \vdots \\ a_{Q1} & a_{Q2} & \cdots & a_{QD} \end{bmatrix}.$$

This further imposes $D^2 - \big(D(D-1)/2\big)$ restrictions, which is necessary and sufficient for estimating the model.

2.4 Prior Distribution

This section discusses the prior distributions and model representation used in the paper to draw from the posterior distributions of non-parametric functions, factors and other parameters.

2.4.1 Prior for Non-Parametric Functions

The non-parametric functions g_{qj} are modelled as the second-order Markov process priors. The prior imposes a structure on the functions where deviations from linearity are penalized.

The prior is given as

$$g_{qjm} = \Big(1 + \frac{h_{qjm}}{h_{qj,m-1}}\Big) g_{qj,m-1} - \Big(\frac{h_{qjm}}{h_{qj,m-1}}\Big) g_{qj,m-2} + u_{qjm}, \tag{2.11}$$

where $h_{qjm} = v_{qjm} - v_{qj,m-1}$ (m is the index for unique ordered value) and $u_{qjm} \sim N(0, \tau_{qj}^2 h_{qjm})$. The prior shrinks the functions to take a linear form and deviations are allowed if enough evidence is present in the data. This can be seen by rewriting the second order prior as

$$g_{qjm} = g_{qj,m-1} + \frac{h_{qjm}}{h_{qj,m-1}}\Big(g_{qj,m-1} - g_{qj,m-2}\Big) + u_{qjm}, \tag{2.12}$$

The slope value of the function at $(m-1)^{th}$ value is extended to the m^{th} value. The mean slope for the function is the same for the consecutive order values and deviations are allowed via u_{qjm} whose variance is linearly proportional to distance h_{qjm}. As discussed, small values of τ produce smoother functions. Distribution of the initial state of g_{qj} is necessary for the Markov prior to being proper. The initial state can be modelled as

$$\begin{pmatrix} g_{qj1} \\ g_{qj2} \end{pmatrix} \sim N\left(\begin{pmatrix} g_{qj10} \\ g_{qj20} \end{pmatrix}, \tau_{qj}^2 G_{qj0} \right). \tag{2.13}$$

The prior for g can be written as $H_{qj} g_{qj} = u_{qj}$ where $u_{qj} \sim N(\hat{\mu}_{qj}, \Sigma_{qj})$ being the error term

in the Markov process with $\hat{\mu}_{qj} = (g_{qj10}, g_{qj20}, 0, \ldots, 0)'$ and H_{qj} and Σ_{qj} defined as

$$H_{qj} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \frac{h_{qj3}}{h_{qj2}} & -\left(1+\frac{h_{qj3}}{h_{qj2}}\right) & 1 & \ddots & \vdots \\ & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \frac{h_{qjm_j}}{h_{qj(m_j-1)}} & -\left(1+\frac{h_{qjm_j}}{h_{qj(m_j-1)}}\right) & 1 \end{bmatrix} \quad and$$

$$\Sigma_{qj} = \begin{bmatrix} G_{qj0} & 0 & \cdots & 0 \\ 0 & h_{qj3} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{qjm_j} \end{bmatrix}.$$

A simple change in variable from u_{qj} to g_{qj} will give the conditional distribution as $g_{qj}|\tau^2_{qj} \sim N(g_{qj0}, \tau^2_{qj}K^{-1}_{qj})$ where the penalty matrix $K_{qj} = H'_{qj}\Sigma^{-1}_{qj}H_{qj}$ and $g_{qj0} = H^{-1}_{qj}\hat{\mu}_{qj}$.

2.4.2 Prior for Factor

The representation in Eq. 2.7 is made even more compact by stacking the observation equations over each other to enable the drawing of f from its joint conditional distribution.

$$\begin{aligned} y &= Q_1 g_1 + M_{00}Q_2 g_2 + \cdots + M_{00}Q_J g_J + Af + \epsilon \\ H_f f &= v, \end{aligned} \tag{2.14}$$

where,

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_Q \end{bmatrix}, \; Q_j = \begin{bmatrix} Q_{1j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_{Qj} \end{bmatrix}, \; g_j = \begin{bmatrix} g_{1j} \\ \vdots \\ g_{Qj} \end{bmatrix}, \; A = \begin{bmatrix} A_1 \\ \vdots \\ A_Q \end{bmatrix}, \; \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_Q \end{bmatrix},$$

$$\Omega_1^* = \begin{bmatrix} I_T \times \omega_{11} & I_T \times \omega_{12} & \cdots & I_T \times \omega_{1Q} \\ I_T \times \omega_{21} & I_T \times \omega_{22} & \cdots & I_T \times \omega_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ I_T \times \omega_{Q1} & I_T \times \omega_{Q2} & \cdots & I_T \times \omega_{QQ} \end{bmatrix}, \; M_{00} \; = \; \begin{bmatrix} M_0 & 0 & \cdots & 0 \\ 0 & M_0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & M_0 \end{bmatrix}$$

and $\epsilon \sim N(0, \Omega_1^*)$. The VAR equation is considered as the likelihood function. A simple change in variable from v to f in the dynamic factor equation will lead to $f \sim N(0, K_f^{-1})$ where the precision matrix $K_f = H_f' S^{-1} H_f$. Initial states for all the factors are defined as $f_{d1} \sim N(0, \sigma_d^2/(1-\gamma_d^2))$. The initial conditions ensure that the prior distribution is proper. Therefore, the matrix $\Omega_S = diag\{\sigma_1^2/(1-\gamma_1^2), \ldots, \sigma_D^2/(1\gamma_D^2)\}$.

2.4.3 Prior for other parameters

The priors for other coefficients and hyperparameters for $q = 1, \ldots, Q$, $j = 1, \ldots, J$ and $d = 1, \ldots, D$ are given as

$$\begin{aligned} \tau_{qj}^2 &\sim IG(\frac{\nu_{qj0}}{2}, \frac{V_{qj0}}{2}) \\ \Omega_1^{-1} &\sim W(r_0, R_0) \\ a &\sim N(a_0, A_0) \\ \gamma_d &\sim TN_{(-1,1)}(\gamma_{d0}, G_{d0}) \\ \sigma_d^2 &\sim IG(\frac{\sigma_{d0}}{2}, \frac{S_{d0}}{2}), \end{aligned} \tag{2.15}$$

where $a = \{a_1', \ldots, a_Q'\}'$ is a $QD \times 1$ dimensional vector. The new vector representation of loadings will be useful for specifying its full conditional distribution.

2.5 Estimation

The paper first considers estimating a univariate Gaussian non-parametric model with one dynamic factor and one lag to motivate the estimation. Let y_t and f_t be scalar quantities. The model will take the following form

$$\begin{aligned} y_t &= g_t(y_{t-1}) + a f_t + \epsilon_t \\ f_t &= \gamma f_{t-1} + v_t. \end{aligned} \tag{2.16}$$

The stacked version of the univariate model can be written as

$$\begin{aligned} y &= Qg + Af + \epsilon\,, \qquad \epsilon \sim N(0, \sigma_\epsilon^2 I) \\ H_f f &= v\,, \qquad\qquad\quad v \sim N(0, \sigma_v^2 I). \end{aligned} \tag{2.17}$$

Let θ denote the set of all model parameters (the vector will change depending upon the context in the following sections). The MCMC estimation for the univariate model involves sequential sampling from the conditional distributions. However, sequential sampling of factors and their loadings suffers from the problem of slow convergence and poor mixing. To avoid this, the paper considers sampling factors and their loading jointly through the block sampling scheme (Chib and Jeliazkov (2006), Chan and Jeliazkov (2009)). The following steps state the estimation algorithm.

Algorithm 1: MCMC Implementation of the Univariate Gaussian Non-Parametric Dynamic Factor Model

STEP 1. Sample $[g|y, \theta_{\{-g\}}] \sim N(\hat{g}, \hat{G})$ where $\hat{G} = (\frac{1}{\tau^2}K + \frac{1}{\sigma_\epsilon^2}Q'Q)^{-1}$ and $\hat{g} = \hat{G}\big(\frac{1}{\tau^2}Kg_0 + \frac{1}{\sigma_\epsilon^2}Q'(y - Af)\big)$ where g_0 is the mean of the prior for g and $K = K_{qj}$ as defined earlier.

STEP 2. Sample $[\tau^2|g] \sim IG\Big((\nu_0 + m)/2, \big(V_0 + (g - g_0)'K(g - g_0)\big)/2\Big)$ where ν_0 and V_0 are parameters for the inverse gamma prior for τ^2 and m is the count of unique elements of y.

STEP 3. Sample $[\sigma_\epsilon^2|y, \theta_{\{-\sigma_\epsilon^2\}}] \sim IG\Big((r_0 + T)/2, \big(R_0 + (y - Qg - Af)'(y - Qg - Af)\big)/2\Big)$ where r_0 and R_0 are parameters for the inverse gamma prior for σ_ϵ^2.

STEP 4. Sample $[a, f|y, \theta_{-\{a,f\}}]$ in one block as

1. Sample $[a|y, \theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with

tailored proposal $a^* \sim q(\tilde{a}, W)$ and accept a^* with probability

$$\alpha_{MH}(a, a^*) = min\left\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a}, W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a}, W)}\right\}.$$

2. Sample $[f|y, \theta_{-\{f\}}] \sim N(\hat{f}, \hat{F})$ where $\hat{F} = \left(K_f + \frac{1}{\sigma_v^2}A'A\right)^{-1}$ and $\hat{f} = \hat{F}\big(K_f\tilde{f} + \frac{1}{\sigma_v^2}A'(y - Qg)\big)$.

STEP 5. Sample $[\gamma|f, \sigma_v^2]$ by MH algorithm with proposal $\gamma^* \sim N(\hat{\gamma}, \hat{G})$ where $\hat{G} = (G_0^{-1} + (f'_{1:T-1}f_{1:T-1})/\sigma_v^2)^{-1}$ and $\hat{\gamma} = \hat{G}(G_0^{-1}\gamma_0 + (f'_{1:T-1}f_{2:T})/\sigma_v^2)$. The proposal γ^* is accepted with the probability

$$\alpha_{MH}(\gamma, \gamma^*) = min\left\{1, \frac{f_N(f_1|0, \sigma_v^2/(1-\gamma^{*2}))}{f_N(f_1|0, \sigma_v^2/(1-\gamma^2))}\right\}.$$

STEP 6. Sample $[\sigma_v^2|f, \gamma] \sim IG\Big(\sigma_0 + T/2, \big(S_0 + (f^* - \bar{f})'(f^* - \bar{f})\big)/2\Big)$ where $f^* = (f_1\sqrt{1-\gamma^2}, f_2, \ldots, f_T)$ and $\bar{f} = (0, \gamma f_2, \ldots, \gamma f_T)$.

The tailored proposal density $a^* \sim q(\tilde{a}, W)$ used in Step 4 is suggested to be a multivariate Student's t distribution with low degrees of freedom to ensure heavy tails with $\hat{a}$ and W being the mode and inverse of the negative Hessian at the mode of $[a|y, \theta_{-f}]$ (Chib (1996)). The conditional density of a marginal of f is obtained as

$$\begin{aligned}\pi(a|y, \theta_{\{-f,a\}}) &= \frac{\pi(a|y, \theta_{\{-f\}})\pi(f|y, \theta_{\{-f,a\}})}{\pi(f|y, \theta_{\{-a\}})} \\ \pi(a|y, \theta_{\{-f,a\}}) &\propto \frac{\pi(a|y, \theta_{\{-f\}})}{\pi(f|y, \theta_{\{-a\}})},\end{aligned} \tag{2.18}$$

which is not difficult to obtain since full conditional densities for a and f are known in the model and $\pi(f|y, \theta_{\{-a\}})$ is absorbed in the constant of proportionality. The sampling of g in step 1 involves the inversion of an $m \times m$ matrix. To avoid inverting using brute force,

Remark 1 explains a trick which leads to computational efficiency.

Remark 1: Sampling of Non-Centered Functions. Since the matrices K and $Q'Q$ are banded, the inverse of the variance-covariance matrix, $\hat{G}^{-1}$, is banded as well. Computational efficiency can be achieved by avoiding the inversion of $(\tau^2 K + \frac{1}{\sigma_\epsilon^2} Q'Q)$. In order to obtain a random draw from $N(\hat{g}, \hat{G})$ efficiently, first sample $u \sim N(0, I)$, and then solve $Pw = u$ for w by back substitution where P is the Cholesky decomposition of $\hat{G}^{-1}$, also banded. It follows that $w \sim N(0, \hat{G})$. Adding the mean $\hat{g}$ to w, will be equivalent to drawing $g \sim N(\hat{g}, \hat{G})$. The mean $\hat{g}$ is found by solving $\hat{G}^{-1}\hat{g} = \frac{1}{\tau^2} K g_0 + \frac{1}{\sigma_\epsilon^2} Q'(y - Af)$, which is done in $O(T)$ operations by back substitution.

Turning attention to the DF-NPVAR, given the identification restrictions in Sec. 2.3 and the priors for the parameters in Sec. 2.4, the MCMC estimation can proceed through an iterative sampling of the following steps.

Algorithm 2: MCMC Implementation of the DF-NPVAR Model

STEP 1. Sample the first function (non-centered) in q^{th} equation as $[g_{q1}|y, \theta_{\{-g_{q1}\}}] \sim N(\hat{g}_{q1}, \hat{G}_{q1})$ where

$$\begin{aligned} \hat{G}_{q1} &= \Big(\frac{1}{\tau_{q1}^2} K_{q1} + \frac{1}{\omega_{q|\{-q\}}} Q_{q1}' Q_{q1}\Big)^{-1} \\ \hat{g}_{q1} &= \hat{G}_{q1}\Big(\frac{1}{\tau_{q1}^2} K_{q1} g_{q10} + \frac{1}{\omega_{q|\{-q\}}} Q_{q1}'\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj} g_{qj} - A_q f\big)\Big) \end{aligned}$$

with $\mu_{q|\{-q\}} = E(\epsilon_q|\epsilon_{-q})$ and $\omega_{q|\{-q\}} = Var(\epsilon_q|\epsilon_{-q})$. The centered functions are sampled as $[g_{qj}|y, \theta_{\{-g_{qj}\}}] \sim N(\hat{g}_{q1}, \hat{G}_{q1})$ for $j = 2, \ldots, J$ where

$$
\begin{aligned}
\hat{G}_{qj} &= \Big(\frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}M_0Q_{qj}\Big)^{-1} \\
\hat{g}_{qj} &= \hat{G}_{qj}\Big(\frac{1}{\tau_{qj}^2}K_{qj}g_{qj0} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}M_0\big(y_q - \mu_{q|\{-q\}} - \sum_{r\geq 2, k\neq j} M_0Q_{qr}g_{qr} \\
&\quad -Q_{q1}g_{q1} - A_qf\big)\Big).
\end{aligned}
$$

STEP 2. Sample $[\tau_{qj}^2|g_{qj}] \sim IG\Big((\nu_{qj0}+m_j)/2, \big(V_{qj0} + (g_{qj}-g_{qj0})'K_{qj}(g_{qj}-g_{qj0})\big)/2\Big)$.

STEP 3. Sample $[\Omega_1^{-1}|y, \theta_{\{-\Omega_1\}}] \sim W\Big(r_0+T, \big(R_0^{-1} + \sum_{t=1}^{T} e_te_t'\big)^{-1}\Big)$ where e_t denotes Q vector of residuals in time period t.

STEP 4. Sample $[a, f|y, \theta_{-\{a,f\}}]$ in one block as

1. Sample $[a|y, \theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with tailored proposal $a^* \sim q(\tilde{a}, W)$ and accept a^* with probability

$$
\alpha_{MH}(a, a^*) = min\left\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a}, W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a}, W)}\right\}.
$$

2. Sample $[f|y, \theta_{-\{f\}}] \sim N(\hat{f}, \hat{F})$ where

$$
\begin{aligned}
\hat{F} &= \big(K_f + A'(\Omega_1^*)^{-1}A\big)^{-1} \\
\hat{f} &= \hat{F}\Big(K_f\tilde{f} + A'(\Omega_1^*)^{-1}\big(y - Q_1g_1 - \sum_{j=2}^{J} M_{00}Q_Jg_J\big)\Big).
\end{aligned}
$$

STEP 5. Sample $[\gamma_d|f_d, \sigma_d^2]$ by MH algorithm with proposal $\gamma_d^* \sim N(\hat{\gamma}_d, \hat{G}_d)$ where $\hat{G}_d = (G_{d0}^{-1} + (f'_{\{d,1:T-1\}}f_{\{d,1:T-1\}})/\sigma_d^2)^{-1}$ and $\hat{\gamma}_d = \hat{G}_d\big(G_{d0}^{-1}\gamma_{d0} + (f'_{\{d,1:T-1\}}f_{\{d,2:T\}}/\sigma_d^2)\big)$ where $f_{\{d,n:m\}}$ is the vector of d^{th} factor from time n to m. The proposal γ_d^* is accepted with

the probability

$$\alpha_{MH}(\gamma_d, \gamma_d^*) = min\left\{1, \frac{\phi(f_1|0, \frac{\sigma_d^2}{1-\gamma_d^{*2}})}{\phi(f_1|0, \frac{\sigma_d^2}{1-\gamma_d^2})}\right\},$$

where $\phi(.)$ is the Gaussian density.

STEP 6. Sample $[\sigma_d^2|f_d, \gamma_d] \sim IG\Big((\sigma_{d0}+T)/2, \big(S_{d0} + (f_d^* - \bar{f}_d)'(f_d^* - \bar{f}_d)\big)/2\Big)$ where f_d^* $= (f_{k1}\sqrt{1-\gamma_d^2}\ , f_{k2}, \ldots, f_{kT})'$ and $\bar{f}_d = (0, \gamma_d f_{k2}, \ldots, \gamma_d f_{kT})'$.

A slight change in representation of Eq. 2.14 is required to obtain $\pi(a|y, \theta_{\{-f\}})$ which is essential to obtain $\pi(a|y, \theta_{\{-f,a\}})$ in Eq. 2.18.

$$y = Q_1 g_1 + M_{00}Q_2 g_2 + \cdots + M_{00}Q_J g_J + F_f a + \epsilon, \tag{2.19}$$

where $F_f = diag\{\tilde{F}, \ldots, \tilde{F}\}$ is a $TQ \times TD$ dimensional vector with $\tilde{F} = \{f_1', \ldots, f_T'\}'$. Thus, the full conditional density of a will take the form $[a|y, \theta_{-\{a\}}] \sim N(\hat{a}, \hat{A})$ where

$$\begin{aligned}
\hat{A} &= \big(A_0^{-1} + F_f'(\Omega_1^*)^{-1}F_f\big)^{-1} \\
\hat{a} &= \hat{A}\Big(A_0^{-1}a_0 + F_f'(\Omega_1^*)^{-1}\big(y - Q_1 g_1 - \sum_{j=2}^{J} M_{00}Q_J g_J\big)\Big).
\end{aligned}$$

The trick in *Remark 1* will not work on centred functions in step 1 since $Q_{qj}'M_0Q_{qj}$ is not banded. As shown in Jeliazkov and Lee (2010), an application of Sherman-Morrison formulae will ease the computational cost, which is explained in *Remark 2*.

Remark 2: Sampling of Centered Functions. To avoid inverting $\hat{G}_{qj}^{-1}$, one can use the defini-

tion of residual maker matrix M_0 inside $\hat{G}_{qj}$ which is given as

$$\hat{G}_{qj} = \Big(\frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}Q_{qj} - \frac{1}{\omega_{q|\{-q\}}T}c'_{qj}c_{qj}\Big)^{-1},$$

where $c_{qj} = Q'_{qj}l$. Let $A_{qj} = \frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}Q_{qj}$, $u_{qj} = \frac{c_{qj}}{\sqrt{\omega_{q|\{-q\}}T}}$ and $\lambda_{qj} = u'_{qj}A_{qj}u_{qj}$. Using the Sherman-Morrison formulae, $\hat{G}_{qj}$ can be written as

$$\hat{G}_{qj} \;=\; A_{qj}^{-1} + \frac{A_{qj}^{-1}u_{qj}u'_{qj}A_{qj}^{-1}}{1-\lambda_{qj}}.$$

Efficiency gains are achieved as $\hat{g}_{qj}$ can be obtained without inverting A_{qj}. Let $B_{qj} = (A_{qj} + \frac{u_{qj}u'_{qj}}{1-\lambda_{qj}})$, and thus $\hat{G}_{qj} = A_{qj}^{-1}B_{qj}A_{qj}^{-1}$. Following the below-mentioned steps, the sampling of g_{qj} from $N(\hat{g}_{qj}, \hat{G}_{qj})$ can be done through $O(T)$ operations instead of $O(T^3)$.

STEP 1. Draw $w_1 \sim N(0, A_{qj})$ and $w_2 \sim N(0, 1)$.

STEP 2. Let $w_3 = w_1 + w_2 u_{qj}\sqrt{1-\lambda_{qj}}$ so that $w_3 \sim N(0, B_{qj})$.

STEP 3. Let $w_4 = A_{qj}w_3$ so that $w_4 \sim N(0, \hat{G}_{qj})$.

STEP 4. Let $g_{qj} = \hat{g}_{qj} + w_4$ so that $g_{qj} \sim N(\hat{g}_{qj}, \hat{G}_{qj})$.

2.6 Model Comparison

Since the researcher's take on a specific problem is reflected in how a model is specified and estimated, it is important to know which model accounts for uncertainty properly. Given models M_i and M_k, Bayesian formulation provides a straightforward way to compare models

using posterior odds, which involves the prior odds and ratio of marginal likelihoods (Bayes factor). The posterior odds take the following form

$$\frac{p(M_i|y)}{p(M_k|y)} = \frac{p(M_i)\, m(y|M_i)}{p(M_k)\, m(y|M_k)}.$$

The marginal likelihood $m(y|M_i)$ is defined as

$$m(y|M_i) = \int p(y|\theta_i, M_i)p(\theta_i|M_i)d\theta_i,$$

where $p(y|M_i, \theta_i)$ is the likelihood for M_i and $p(\theta_i|M_i)$ is the priors on the parameter vector θ_i used in M_i. The additive function framework used in the DF-NPVAR model makes θ_i high dimensional, making numerical integration highly costly. Using the application of Bayes theorem, Chib (1995) suggested a more tractable formula to calculate the marginal likelihood, which is given as

$$m(y|M_i) = \frac{p(y|\theta_i^*, M_i)p(\theta_i^*|M_i)}{p(\theta_i^*|y, M_i)},$$

where θ_i^* is the posterior mean. In the case of the DF-NPVAR model, the marginal likelihood will be calculated as

$$m(y) = \frac{p(y|g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*})p(g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*})}{p(g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}|y)}$$

given the Gaussian structure of the model, the marginal likelihood can be calculated as

$$m(y) = \frac{p(y|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}) p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*})}{p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}|y)}, \tag{2.20}$$

where densities are marginalized over g. This is possible since conditional on other parameters, the density $p(y|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*})$ is still normal (Koop and Poirier (2004)). This helps avoid evaluating the densities at high dimensional g, saving computational costs. The numerator in Eq. 2.20 is readily available as defined in this paper. To evaluate the posterior density at the posterior mean value, the paper uses the law of probability as

$$\begin{aligned} p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}|y) = p(\tau^{2*}|y) p(\Omega_1^*|\tau^{2*}, y) p(a^*|\Omega_1^*, \tau^{2*}, y) \dots \\ p(\sigma^{2*}|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, y). \end{aligned} \tag{2.21}$$

Chib (1995) provided a method of calculating marginal likelihood under Gibbs sampling. Since the full conditionals of τ^2, Ω_1, f, a and σ^2 are known, the marginal densities in Eq. 2.21 can be estimated using Rao-Blackwellization (Tanner and Wong (1987); Gelfand and Smith (1990)). For given parameters θ_i and θ_k, the conditional density of θ_i marginal of θ_k can be calculated as

$$p(\theta_i|y) = \frac{1}{B} \sum_{b=1}^{B} p(\theta_i|\theta_k^{\{b\}}, y)$$

using B draws of θ_k from the MCMC run. For the estimation of the marginal density of γ, one can refer to Chib and Jeliazkov (2001, 2005).

2.7 Model Extensions

The estimation framework for the DF-NPVAR model is flexible and readily applicable to various situations. This is possible since the estimation of non-parametric functions and factors can be done conditionally on the adjustments done in other parts of the model. This section illustrates two such model extensions which further DF-NPVAR model applicability in familiar settings and support the claim for the model's modularity.

2.7.1 DF-NPVAR with Serially Correlated Errors

The model in Eq.2.7 is extended by considering the serially correlated error term that follows a mean zero stationary $AR(N)$ process.

$$\epsilon_{qt} = \rho_1 \epsilon_{q,t-1} + \cdots + \rho_N \epsilon_{q,t-N} + \eta_{qt}, \tag{2.22}$$

where $\eta_{qt} \sim N(0,1)$. Let $\epsilon_q = (\epsilon_{q1}, \ldots, \epsilon_{qT})'$ where $\epsilon_q \sim N(0, \Omega_q)$. Given the AR(N) structure of ϵ_{qt}, Ω_q is a $T \times T$ dimensional Toeplitz matrix. Generally, Ω_q can be determined in the following way. Let $\varphi_n = E(\epsilon_{qt}\epsilon_{q,t-n})$ be the n^{th} autocovariance term of Ω_q. Due to symmetry, $\varphi_n = \varphi_{-n}$. It can be shown that the autocovariance itself follow an AR(N) process as $\varphi_n = \rho_1 \varphi_{n-1} + \cdots + \rho_N \varphi_{n-N}$. The first N values $(\varphi_0, \ldots, \varphi_{N-1})$ are given by the first N elements of the first column of the $N^2 \times N^2$ matrix $[I - F \otimes F]^{-1}$ where F, an $N \times N$

matrix, is given as

$$F = \begin{bmatrix} & \rho' & \\ & & \\ I_{N-1} & & 0_{(N-1)\times 1} \end{bmatrix}.$$

Thus, Ω_q can be obtained as $\Omega_q[i,k] = \varphi_{i-k}$. To illustrate, if the process considered is AR(1) process, $\Omega_q[i,k] = \rho^{|i-k|}/(1-\rho^2)$.

The prior density for ρ is defined as $\rho \sim N(\rho_0, P_0)I_{S_\rho}$ where $\rho = (\rho_1, \ldots, \rho_N)'$, I_{S_ρ} is the indicator function which takes the value 1 in case ρ belongs to the set of ρ vectors which satisfy stationarity. Due to the new structure of the variance-covariance matrix, *Remark 2* cannot be implemented, making computing its inverse a very time-consuming task. Following Chib and Jeliazkov (2006), the error terms can be orthogonalized by breaking the variance-covariance matrix into correlated and uncorrelated parts. Then, the correlated part is treated as random effects, and the model is tweaked to take the homoskedastic form. To illustrate, lets decompose $\Omega_q = \Sigma_q - \kappa I$ where Σ_q is a symmetric positive definite matrix and $\kappa > 0$. It is assumed that κ is equal to the minimum of the set of eigenvalues of Ω_q, which ensures that Ω_q can be decomposed. Furthermore, let $C_q'C_q = \Sigma_q$ where C_q be the Cholesky decomposition of Σ_q and $\Omega_q = C_q'C_q - \kappa I$. The model in Eq.2.7 can be rewritten as

$$\begin{aligned} y_1 &= Q_{11}g_{11} + Q_{12}g_{12} + \cdots + Q_{1J}g_{1J} + A_1 f + C_1' u_1 + \xi_1 \\ &\vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ y_Q &= Q_{Q1}g_{Q1} + Q_{Q2}g_{Q1} + \cdots + Q_{QJ}g_{QJ} + A_Q f + C_Q' u_Q + \xi_Q \\ H_f f &= v. \end{aligned} \tag{2.23}$$

where $u_q \sim N(0, I)$ for all q. Stacking over q, the model can be compactly written as

$$\begin{aligned} y &= Q_1 g_1 + Q_2 g_2 + \cdots + Q_J g_J + Af + C'u + \xi\,, \quad \xi \sim N(0, \kappa I) \\ H_f f &= v, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad v \sim N(0, S), \end{aligned} \tag{2.24}$$

where $u = \{u_1', \ldots, u_Q'\}'$ and

$$C = \begin{bmatrix} C_1 & & \\ & \ddots & \\ & & C_Q \end{bmatrix}.$$

Given $C'u$, which is computationally inexpensive to calculate, the variance is constant across time since $var(\xi) = \kappa I$. The MCMC algorithm to draw from the posterior distribution will take the following form.

Algorithm 3: MCMC Implementation of the DF-NPVAR Model with Serially-Correlated Errors

STEP 1. For $q = 1, \ldots, Q$

1. Sample $[u_q | y_q, \theta_{\{-u_q\}}] \sim N(\hat{u}_q, \hat{U}_q)$ where, $\hat{U}_q = (I + C_q C_q'/\kappa)^{-1}$ and $\hat{u}_q = \hat{U}_q C_q (y_q - \mu_{q|\{-q\}} - Q_{q1} g_{q1} - \sum_{r \geq 2} M_0 Q_{qr} g_{qr} - A_q f)/\kappa$.

2. Sample the first function in equation q as $[g_{q1}|y, \theta_{\{-g_{q1}\}}] \sim N(\hat{g}_{q1}, \hat{G}_{q1})$ where

$$\begin{aligned}
\hat{G}_{q1} &= \Big(\frac{1}{\tau_{q1}^2}K_{q1} + \frac{1}{\kappa}Q_{q1}'Q_{q1}\Big)^{-1} \\
\hat{g}_{q1} &= \hat{G}_{q1}\Big(\frac{1}{\tau_{q1}^2}K_{q1}g_{q10} + \frac{1}{\kappa}Q_{q1}'\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj} g_{qj} - A_q f \\
&\quad -C_q u_q\big)\Big).
\end{aligned}$$

3. The centered functions are sampled as $[g_{qj}|y, \theta_{\{-g_{qj}\}}] \sim N(\hat{g}_{qj}, \hat{G}_{qj})$ for $j = 2, \ldots, J$ where,

$$\begin{aligned}
\hat{G}_{qj} &= \Big(\frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\kappa}Q_{qj}'M_0Q_{qj}\Big)^{-1} \\
\hat{g}_{qj} &= \hat{G}_{qj}\Big(\frac{1}{\tau_{qj}^2}K_{qj}g_{qj0} + \frac{1}{\kappa}Q_{q1}'M_0\big(y_q - \mu_{q|\{-q\}} - \sum_{r\geq 2, k\neq j} M_0 Q_{qr} g_{qr} - \\
&\quad Q_{q1}g_{q1} - A_q f - C_q u_q\big)\Big).
\end{aligned}$$

.

Step 2. Sample $[\tau_{qj}^2|g_{qj}]$ as in Algorithm 2.

Step 3. Sample $[a, f|y, \theta_{-\{a,f\}}]$ in one block as

1. Sample $[a|y, \theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with tailored proposal $a^* \sim q(\tilde{a}, W)$ and accept a^* with probability

$$\alpha_{MH}(a, a^*) = min\left\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a}, W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a}, W)}\right\}.$$

2. Sample $[f|y, \theta_{-\{f\}}] \sim N(\hat{f}, \hat{F})$ where

$$\begin{aligned}
\hat{F} &= \big(K_f + \frac{A'A}{\kappa}\big)^{-1} \\
\hat{f} &= \hat{F}\Big(K_f\tilde{f} + \frac{A'}{\kappa}\big(y - Q_1 g_1 - \sum_{j=2}^{J} M_{00} Q_J g_J - C' u\big)\Big).
\end{aligned}$$

Step 4. Sample $[\gamma_d|f_k, \sigma_k^2]$ as in Algorithm 2.

STEP 5. Sample $[\sigma_d^2|f_k, \gamma_k]$ as in Algorithm 2.

STEP 6. Sample $[\rho|y, \theta_{\{-\rho\}}] \propto \phi(\rho)N(\hat{\rho}, P)I_{S_\rho}$ where $\hat{\rho} = P(P_0^{-1}\rho_0 + E'e)$, $P = (P_0^{-1} + E'E)^{-1}$, and $\phi(\rho) = |\Omega_\rho|^{-Q/2}exp\{-\frac{1}{2}\sum_{q=1}^{Q} e_{q1}'\Omega_\rho^{-1}e_{q1}\}$. Using Metropolis-Hasting algorithm, the proposal (say ρ^*) is drawn from $N(\hat{\rho}, P)I_{S_\rho}$ (Chib and Greenberg (1994)) and accepted with the probability

$$\alpha_{MH}(\rho, \rho^*) = min\Big\{1, \frac{\phi(\rho^*)}{\phi(\rho)}\Big\}$$

where, $e_{qt} = y_{qt} - g_{q1}(s_{q1t}) - \cdots - g_{qJ}(s_{qJt}) - a_q' f_t$, $e_q = (e_{q,N+1}, \ldots, e_{q,T})'$, $e = (e_1', \ldots, e_Q')'$ and E denote the $(T-N)Q \times N$ matrix with rows containing N lags of $e_{qt} = (e_{q,t-1}, \ldots, e_{q,t-N})$ and $t \geq N+1$. The initial N values of e_{qt} corresponding to each equation is $e_{q1} = (e_{q1}, \ldots, e_{qN})'$. The matrix Ω_ρ is $N \times N$ stationary covariance matrix constructed exactly as Ω_q.

2.7.2 DF-NPVAR with Binary Variable

Binary, or qualitative, variables have been extensively utilized in Vector Autoregression (VAR) models. A prevalent example of such a variable is a dummy representing business cycle recessions and expansions. Dueker (2005) pioneered the introduction of a Markov Chain Monte Carlo (MCMC) technique for estimating qualitative variables within the VAR framework as dependent variables rather than merely as controls. This paper extends the versatility of the DF-NPVAR model by incorporating binary variables. Denote the binary variable as d_t and the latent variable as y_t^*. Although this section focuses on a single binary variable for simplicity, the methodology can be expanded to include multiple binary variables.

The relationship between the binary and latent variables is modelled as follows:

$$\begin{aligned} d_t &= 1 \quad if \quad y_t^* > 0 \\ d_t &= 0 \quad if \quad y_t^* \leq 0. \end{aligned} \tag{2.25}$$

Considering Eq. 2.2, the latent variable will enter the model as

$$\begin{aligned} y_{1t} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{1qpt} + a_1' f_t + \sum_{p=1}^{P} b_{1p} y_{t-p}^* + \epsilon_{1t} \\ \vdots \qquad & \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ y_{Qt} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{Qqpt} + a_Q' f_t + \sum_{p=1}^{P} b_{Qp} y_{t-p}^* + \epsilon_{Qt} \\ y_t^* &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{Q+1,qpt} + a_{Q+1}' f_t + \sum_{p=1}^{P} b_{Q+1,p} y_{t-p}^* + \epsilon_{Q+1,t}. \end{aligned} \tag{2.26}$$

Stacking over variables, Eq. 2.26 can be written as

$$Y_t \;=\; \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qpt} + a^* f_t + \sum_{p=1}^{P} b_p y_{t-p}^* + \epsilon_t, \tag{2.27}$$

where,

$$Y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Qt} \\ y_t^* \end{bmatrix},\; g_{qpt} = \begin{bmatrix} g_{1qpt} \\ \vdots \\ g_{Q+1,qpt} \end{bmatrix},\; a^* = \begin{bmatrix} a_1' \\ \vdots \\ a_{Q+1}' \end{bmatrix},\; and \quad b_p = \begin{bmatrix} b_{1p} \\ \vdots \\ b_{Q+1,p} \end{bmatrix}.$$

The full conditional distribution of y_t^* can be obtained similarly to probit models. Since, the variance of $y^* = (y_1^*, \ldots, y_T^*)$ is not identified, the lower right element of $var(\epsilon_t)$ is kept equal to 1 as

$$\Omega_1 = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1Q} & \omega_{1,Q+1} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2Q} & \omega_{2,Q+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{Q1} & \omega_{Q2} & \cdots & \omega_{QQ} & \omega_{Q,Q+1} \\ \omega_{Q+1,1} & \omega_{Q+1,2} & \cdots & \omega_{Q+1,Q} & 1 \end{bmatrix}.$$

The latent variable y_t^* follows a truncated normal distribution.

$$\begin{aligned} &d_t = 1 \quad then \quad y_t^* \sim TN_{[0,\infty]} \\ &d_t = 0 \quad then \quad y_t^* \sim TN_{[-\infty,0]}. \end{aligned} \tag{2.28}$$

Let Y_{-t} be the full vector time series except for time t. To derive the full conditional density $p(y_t^*|Y_{-t}, y_{1t}, \ldots, y_{Qt}, \theta) = p(y_t^*|\theta^*)$, the density associated with the P lags will be exploited to get a conjugate form. The error terms associated with these P densities are as follows

$$\begin{aligned} \epsilon_t &= Y_t - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qpt} - a^* f_t - \sum_{p=1}^{P} b_p y_{t-p}^* \\ \epsilon_{t+1} &= Y_{t+1} - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qp,t+1} - a^* f_{t+1} - \sum_{p=1}^{P} b_p y_{(t+1)-p}^* \\ &\vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ \epsilon_{t+P} &= Y_{t+P} - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qp,t+P} - a^* f_{t+P} - \sum_{p=1}^{P} b_p y_{(t+P)-p}^*. \end{aligned} \tag{2.29}$$

Let ψ_t be the known part of ϵ_t conditional on $[Y_{-t}, y_{1t}, \ldots, y_{Qt}, \theta]$, then

$$\begin{aligned}
\epsilon_t &= Y_t - \psi_t \\
\epsilon_{t+1} &= \psi_{t+1} - b_1 y_t^* \\
&\vdots \\
\epsilon_{t+P} &= \psi_{t+P} - b_P y_t^*.
\end{aligned} \tag{2.30}$$

The conditional density of y_t^* is a function of $(\epsilon_t, \ldots, \epsilon_{t+P})$ and can be written as

$$\begin{aligned}
p(y_t^*|\theta^*) &= f\Big(exp\{-\frac{1}{2}(\epsilon_t'\Omega_1^{-1}\epsilon_t + \epsilon_{t+1}'\Omega_1^{-1}\epsilon_{t+1} + \cdots + \epsilon_{t+P}'\Omega_1^{-1}\epsilon_{t+P})\}\Big) \\
&= f\Big(exp\{-\frac{1}{2}((Y_t - \psi_t)'\Omega_1^{-1}(Y_t - \psi_t) + (\psi_{t+1} - b_1 y_t^*)'\Omega_1^{-1}(\psi_{t+1} - b_1 y_t^*) \\
&\quad + \cdots + (\psi_{t+P} - b_p y_t^*)'\Omega_1^{-1}(\psi_{t+P} - b_p y_t^*)\}\Big).
\end{aligned}$$

Specifically, the density of $\epsilon_{Q+1,t}$ will be used in the conditional density instead of ϵ_t and thus $p(y_t^*|\theta^*)$ can be written as

$$\begin{aligned}
p(y_t^*|\theta^*) = f\Big(&exp\{-\frac{1}{2}((y_t^* - \psi_t^*)'\Omega_{y^*}^{-1}(y_t^* - \psi_t^*) + (\psi_{t+1} - b_1 y_t^*)'\Omega_1^{-1}(\psi_{t+1} - b_1 y_t^*) \\
&+ \cdots + (\psi_{t+P} - b_p y_t^*)'\Omega_1^{-1}(\psi_{t+P} - b_p y_t^*)\}\Big),
\end{aligned}$$

where, $\Omega_{y^*}^{-1} = var(\epsilon_{Q+1,t}|\epsilon_{1,t}, \ldots, \epsilon_{Q,t})$ and $\psi_t^* = E(\epsilon_{Q+1,t}|\epsilon_{1,t}, \ldots, \epsilon_{Q,t})$. Given the normal distribution, obtaining a conditional distribution from joint distribution is straightforward. After collecting all the cross terms, it can be shown that the conditional density for y_t^* will

take the following form

$$[y_t^*|\theta^*] \sim N(C^{-1}D, C^{-1}), \tag{2.31}$$

where, $C = (\Omega_{y^*}^{-1} + b_1'\Omega_1^{-1}b_1 + \cdots b_P'\Omega_1^{-1}b_P)$ and $D = -\Omega_{y^*}^{-1}\psi_t^* + b_1'\Omega_1^{-1}\psi_{t+1} + \cdots + b_P'\Omega_1^{-1}\psi_{t+P}$. The latent variable y_t^* is drawn from the truncated normal distribution using Eq. 2.31 depending upon whether $d_t = 0$ or 1.

The draw of the coefficient b from its conditional distribution can be done by considering the representation in Eq. 2.7 as

$$\begin{aligned}
y_1 &= Q_{11}g_{11} + \cdots + Q_{1J}g_{1J} + A_1 f + \tilde{y}^* b_1 + \epsilon_1 \\
&\vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\
y_Q &= Q_{Q1}g_{Q1} \cdots + Q_{QJ}g_{QJ} + A_Q f + \tilde{y}^* b_Q + \epsilon_Q \\
y^* &= Q_{Q+1,1}g_{Q+1,1} \cdots + Q_{Q+1,J}g_{Q+1,J} + A_{Q+1} f + \tilde{y}^* b_{Q+1} + \epsilon_{Q+1},
\end{aligned} \tag{2.32}$$

where

$$\tilde{y}^* = \begin{bmatrix} y_0^* & \cdots & y_{-(P-1)}^* \\ \vdots & \ddots & \vdots \\ y_{T-1}^* & \cdots & y_{T-P}^* \end{bmatrix} \; and \; b_q = \begin{bmatrix} b_{1q} \\ \vdots \\ b_{Pq} \end{bmatrix}.$$

The conditional distribution for b_q will take a similar form as any OLS coefficient. Given the prior as $p(b_q) = N(b_{0q}, B_{0q})$, the conditional density is given as $[b_q|y, \theta_{-\{b_q\}}] \sim N(\hat{b}_q, \hat{B}_q)$

where for $q = 1, \ldots, Q$

$$\begin{aligned}
\hat{B}_q &= \Big(B_{0q}^{-1} + \frac{1}{\omega_{q|\{-q\}}}\tilde{y}^{*'}\tilde{y}^{*}\Big)^{-1} \\
\hat{b}_q &= \hat{B}_q\Big(B_{0q}^{-1}b_{0q} + \frac{1}{\omega_{q|\{-q\}}}\tilde{y}^{*'}\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj} g_{qj} - Q_{q1}g_{q1} - A_q f\big)\Big)
\end{aligned}$$

and for $q = Q + 1$

$$\begin{aligned}
\hat{B}_{Q+1} &= \Big(B_{0,Q+1}^{-1} + \frac{1}{\omega_{Q+1|\{-(Q+1)\}}}\tilde{y}^{*'}\tilde{y}^{*}\Big)^{-1} \\
\hat{b}_{Q+1} &= \hat{B}_{Q+1}\Big(B_{0,Q+1}^{-1}b_{0,Q+1} + \frac{1}{\omega_{Q+1|\{-(Q+1)\}}}\tilde{y}^{*'}\big(y^{*} - \mu_{Q+1|\{-(Q+1)\}} - Q_{Q+1,1}g_{Q+1,1} - \\
&\qquad \sum_{j=2}^{J} M_0 Q_{Q+1,j} g_{Q+1,j} - A_{Q+1} f\big)\Big).
\end{aligned}$$

2.8 Application to US Macroeconomic Data

The current application contains U.S. post-war quarterly macroeconomic data from 1954:Q1 to 2022:Q4. The set of variables includes output growth (y_t) measured by the percentage change of GDP between two consecutive quarters (seasonally adjusted), average quarterly unemployment rate (u_t), inflation rate (π_t) measured by the percentage change in the Consumer Price Index between consecutive quarters, and interest rates (i_t) measured by the average quarterly market yield on U.S. treasury securities at 10-Year Constant maturity. The data avoids using the secondary market yield on the 3-month Treasury bill as the interest rate since they approached and stayed very close to their lower bound of zero after the great recession. This could lead to findings of nonlinearity due to the effects of the lower

bound, thereby favouring the methods of the paper over a linear model. The variables y_t, u_t and i_t are seasonally adjusted. All four variables are summarized in Tab. 2.1.

Table 2.1: Summary Statistics

Variables	Mean	SD	Min	Max
y_t	6.43	5.07	-29.2	39.7
u_t	5.88	1.64	3.4	13.0
i_t	5.62	2.92	0.65	14.85
π_t	0.80	0.58	-0.40	2.94

DF-NPVAR model is estimated using this data, which helps us examine the behaviour of the dynamic system. The baseline DF-NPVAR(1) model is estimated to achieve parsimony which is given as

$$
\begin{aligned}
y_t &= g_{yy1t}(y_{t-1}) + g_{yu1t}(u_{t-1}) + g_{yi1t}(i_{t-1}) + g_{y\pi 1t}(\pi_{t-1}) + a_y f_t + \epsilon_{yt} \\
u_t &= g_{uy1t}(y_{t-1}) + g_{uu1t}(u_{t-1}) + g_{ui1t}(i_{t-1}) + g_{u\pi 1t}(\pi_{t-1}) + a_u f_t + \epsilon_{ut} \\
i_t &= g_{iy1t}(y_{t-1}) + g_{iu1t}(u_{t-1}) + g_{ii1t}(i_{t-1}) + g_{i\pi 1t}(\pi_{t-1}) + a_i f_t + \epsilon_{it} \\
\pi_t &= g_{\pi y1t}(y_{t-1}) + g_{\pi u1t}(u_{t-1}) + g_{\pi i1t}(i_{t-1}) + g_{\pi\pi 1t}(\pi_{t-1}) + a_\pi f_t + \epsilon_{\pi t} \\
f_t &= \gamma f_{t-1} + v_t\,, \quad \forall \quad t = 1, 2, \ldots, T
\end{aligned}
$$

The model has sixteen non-linear functions and one factor. For the identification strategy on the loadings, the model restricts $a_y = 1$. Since a_y is associated with the output growth equation, the factor itself can be interpreted as the business cycle component in the economy, which affects all the macro variables.

Figure 2.2 presents the estimated functions for the DF-NPVAR(1) model. The Y-axis represents functions and X-axis represents equations. The bold red line is the mean of the non-parametric functions, whereas the light yellow lines represent 95% credible intervals. Many of the relations, especially the own lag effect can be modelled linearly. To a lesser ex-

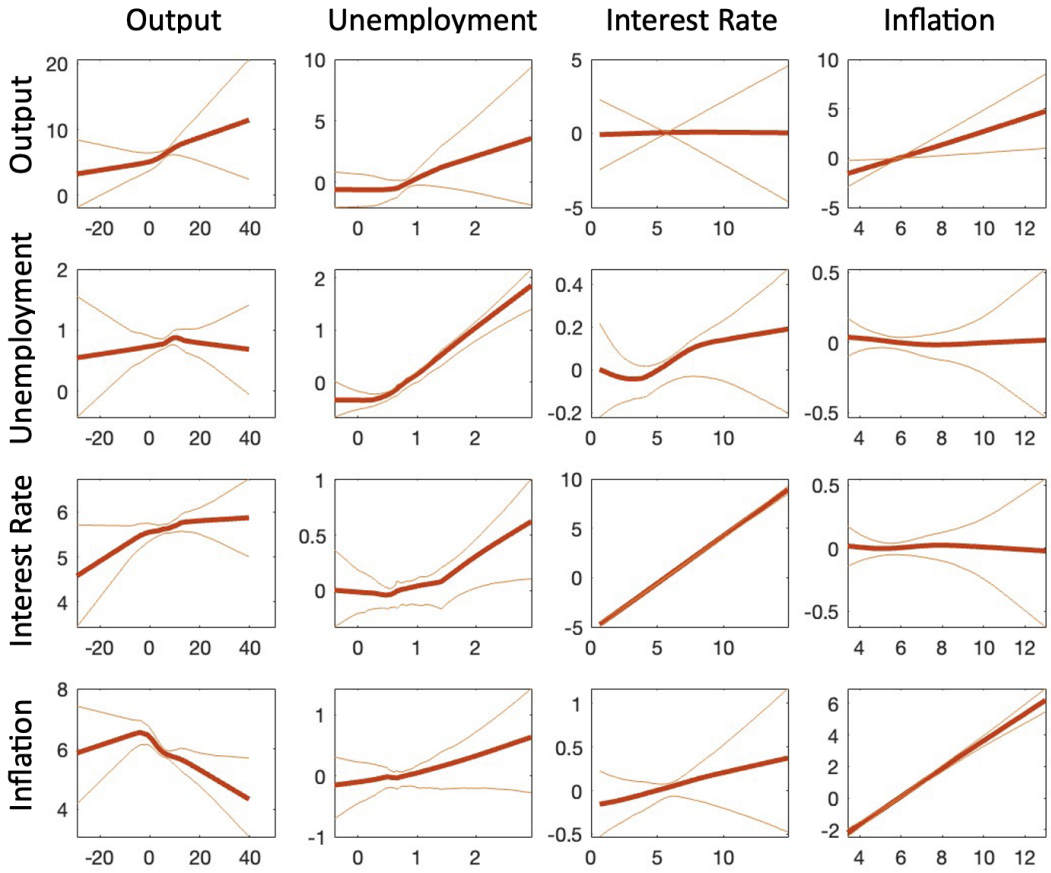


Figure 2.2: Non-parametric functions for the US post-war data

tent, some of the relations can be approximated as linear. For example, the relation between lag interest rates on the output growth rate or lag unemployment on interest rates, as the loss of information due to linearity restriction, is not significant. However, some relations like the effects of lagged inflation on output or lagged unemployment on the interest rate appear to be nonlinear. These results agree with the literature that has found non-linearity in behaviours of output growth (Dahl and Gonzalez-Rivera (03 2);Dahl and González-Rivera (03 1)) and financial markets variables like interest rates and inflation rate Härdle and Tsybakov (1997); Härdle et al. (1998a)). One advantage of using the second-order Markov process prior for g_{qj} is that the linear relationship is preserved if the true relationship is linear. The prior penalizes deviations from linearity, so the model will spit out non-linear relations only

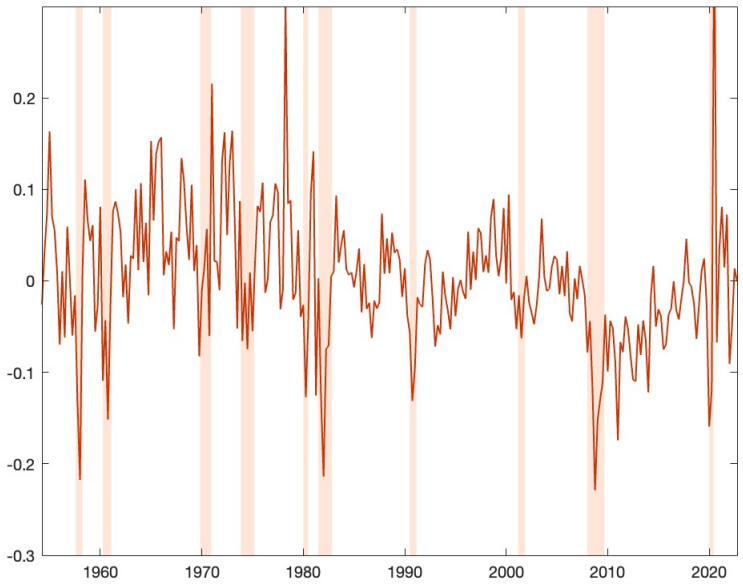


Figure 2.3: Factor representing business cycle for the US post-war data

if enough evidence is there to support it.

The credible intervals are shaped as hourglass since the identification restrictions make the intervals narrower at the point where non-parametric functions are centred. Other identification techniques will have a different effect on the shape of credible intervals. For example, forcing all but one non-parametric function to start from 0 would have led to funnel-shaped credible intervals.

The dynamic factor estimated in this application is shown in Fig 2.3. The factor is superimposed over the shaded periods, representing officially announced recessions by the Fed in the US over the period considered. The factor is able to capture the business cycle element of the economy as it coincides with the recessions. The factor can be used in future research to represent the business cycle component in a macro model.

2.9 Conclusion

This paper has presented the specification, identification, and estimation of non-parametric VAR models with dynamic factors. The paper proposes an efficient MCMC sampling algorithm to estimate the model efficiently. The paper expands on model comparison and extensions on the model. The model is able to accommodate the error terms following AR(1) process, but the idea can be generalized to any form of heteroskedasticity or autocorrelation, which can be taken up in future endeavours. The model's extension with qualitative variables expands its applicability to datasets with binary variables. The application considered U.S. post-war data on GDP growth, unemployment, interest rates, and inflation. The model was able to recover non-linear relationships concerning the financial variables and the factor was able to capture the business cycle component in the economy. Due to the modality of the DF-NPVAR model, it can be extended to consider non-normal error terms and applied to other economies.

Chapter 3

Bayesian Estimation of the Factor Augmented Vector Autoregressive Models

3.1 Introduction

This article revisits the Bayesian estimation of several variants of the factor augmented vector autoregressive (FAVAR) model proposed by Bernanke et al. (2005) (from here on referred to as BBE). Their model represents a key contribution to the parsimonious analysis of high-dimensional intertemporally related outcomes. They employ a factor structure to distil the information content of a large number of economic or financial variables into a small set of latent factors, which, together with key macroeconomic variables, enter a lower-dimensional vector autoregression. The model has gained popularity in macroeconomics and finance (Eickmeier et al. (2014), Caggiano et al. (2014), Boivin et al. (2009), Bianchi et al. (2009), Forni and Gambetti (2010), Moench (2008), Ludvigson and Ng (2009) to name a few), and is

also gaining popularity in other fields (Fan et al. (2011), Fan et al. (2013), and Tsai and Tsay (2010)). Despite their considerable utility, applications and extensions of the FAVAR model have been hindered by significant computational complexity, which has typically required simplifications on either the modelling or estimation front, e.g., by not involving covariates or through the use of sequential plug-in estimators. In particular, we focus on the following baseline specification, where, for $t = P, \ldots, T$, we have

$$\begin{aligned} Z_t &= \tilde{X}_{1t}\delta_1 + \Gamma_m y_t + \Lambda_m f_t + \varepsilon_{1t} \\ \begin{pmatrix} f_t \\ y_t \end{pmatrix} &= \tilde{X}_{2t}\delta_2 + \sum_{p=1}^{P-1} B_p \begin{pmatrix} f_{t-p} \\ y_{t-p} \end{pmatrix} + \varepsilon_{2t}\,, \end{aligned} \tag{3.1}$$

where $Z_t = (Z_{1t}, \ldots, Z_{nt})'$ is an $n \times 1$ vector of observable variables, $f_t = (f_{1t}, \ldots, f_{r_1 t})'$ is an $r_1 \times 1$ vector of unobservable factors, and $y_t = (y_{1t}, \ldots, y_{r_2 t})'$ is an $r_2 \times 1$ vector of observable factors. The control matrices $\tilde{X}_{1t}$ and $\tilde{X}_{2t}$ are $n \times nr_3$ and $(r_1 + r_2) \times (r_1 + r_2)r_4$ dimensional and are defined as

$$\begin{aligned} X_{1t} &= \{X_{11t}, \ldots, X_{1r_3 t}\}', \qquad \tilde{X}_{1t} = X_{1t}' \otimes I_n \\ X_{2t} &= \{X_{21t}, \ldots, X_{2r_4 t}\}', \qquad \tilde{X}_{2t} = X_{2t}' \otimes I_{r_1 + r_2} \end{aligned}$$

where $\otimes$ is a Kronecker product. The vectors δ_1 and δ_2 are $nr_3 \times 1$ and $(r_1 + r_2)r_4 \times 1$ respectively, carrying coefficients for the control variables and are defined as

$$\delta_1 = \begin{pmatrix} \delta_{111} \\ \vdots \\ \delta_{11r_3} \\ \vdots \\ \delta_{1n1} \\ \vdots \\ \delta_{1nr_3} \end{pmatrix}, \ \delta_2 = \begin{pmatrix} \delta_2^f \\ \delta_2^y \end{pmatrix}, \ \delta_2^f = \begin{pmatrix} \delta_{211}^f \\ \vdots \\ \delta_{21r_4}^f \\ \vdots \\ \delta_{2r_11}^f \\ \vdots \\ \delta_{2r_1r_4}^f \end{pmatrix}, \ \delta_2^y = \begin{pmatrix} \delta_{211}^y \\ \vdots \\ \delta_{21r_4}^y \\ \vdots \\ \delta_{2r_21}^y \\ \vdots \\ \delta_{2r_2r_4}^y \end{pmatrix}$$

The error term $\varepsilon_{1t} \sim N(0, \Sigma)$ where $\Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$. Let ϵ_t and v_t be innovations to f_t and y_t, respectively such as $\varepsilon_{2t} = \{e_t, v_t\}'$. The error term $\varepsilon_{2t} \sim N(0, \Omega)$ where

$$\Omega = E(\varepsilon_{2t}\varepsilon_{2t}') = \begin{pmatrix} E(e_te_t') & E(e_tv_t') \\ E(v_te_t') & E(v_tv_t') \end{pmatrix} = \begin{pmatrix} \Omega_{ee} & \Omega_{ev} \\ \Omega_{ve} & \Omega_{vv} \end{pmatrix}$$

where Ω_{ee} is $r_1 \times r_1$, Ω_{ev} is $r_1 \times r_2$, Ω_{ve} is $r_2 \times r_1$ and Ω_{vv} is $r_2 \times r_2$ dimensional matrices. The loading matrices $\Gamma_m = (\Gamma_1, \ldots, \Gamma_n)'$ is $n \times r_2$ and $\Lambda_m = (\Lambda_1, \ldots, \Lambda_n)'$ is $n \times r_1$ where $\Gamma_i = (\Gamma_{i1}, \ldots, \Gamma_{ir_2})$ is $1 \times r_2$ and $\Lambda_i = (\Lambda_{i1}, \ldots, \Lambda_{ir_1})$ is $1 \times r_1$ for all $i = 1, \ldots, n$. The matrix

B_p is $(r_1 + r_2) \times (r_1 + r_2))$ dimensional for all $p = 1, \ldots, P-1$ is defined as

$$B_p = \begin{pmatrix} B_p^{ff} & B_p^{fy} \\ B_p^{yf} & B_p^{yy} \end{pmatrix} = \begin{pmatrix} b_{p11}^{ff} & \cdots & b_{p1r_1}^{ff} & b_{p11}^{fy} & \cdots & b_{p1r_2}^{fy} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{pr_11}^{ff} & \cdots & b_{pr_1r_1}^{ff} & b_{pr_11}^{fy} & \cdots & b_{pr_1r_2}^{fy} \\ \\ b_{p11}^{yf} & \cdots & b_{p1r_1}^{yf} & b_{p11}^{yy} & \cdots & b_{p1r_2}^{yy} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{pr_21}^{yf} & \cdots & b_{pr_2r_1}^{yf} & b_{pr_21}^{yy} & \cdots & b_{pr_2r_2}^{yy} \end{pmatrix}.$$

Let θ be the set of all model parameters. BBE introduced the FAVAR model and proposed two ways of estimation. The first method is a two-step principal component approach where the first step estimates f_t using the principal component analysis (PCA) and the second step estimates θ given the estimated f_t. This approach estimates factors non-parametrically making estimation computationally simple and easy to implement. However, the estimation of f_t does not depend on the model due to the sequential estimation procedure. The second method is a Bayesian one-step likelihood approach where the f_t and θ are estimated jointly using a multi-step Gibbs sampler. This method estimates f_t and θ simultaneously but can be computationally costly.

The two-step PCA method gained much more popularity compared to the one-step Bayesian method due to the ease of its implementation. BBE, in their paper, recommended this approach and stated "the advantages of using the computationally more burdensome Gibbs sampling procedure instead of the two-step method appear to be modest in this application". However, the drawbacks associated with the estimation of f_t independent of the model are greatly undermined. Since the estimation is agnostic to the structure of the model, the VAR portion is completely ignored which is vital for the evolution of f_t. The approach,

not abiding by the model, forces researchers to use auxiliary models to estimate standard errors or conduct forecasting. Ideally, the principal components should be extracted from the residuals of the regression of y_t on Z_t in every iteration as θ is updated.

One of the reasons for the unpopular sentiment against the Gibbs sampler is the burdensome computation procedure and slow convergence in large models. On reevaluating the one-step Bayesian likelihood estimation, it appears that their Gibbs sampler shows slow convergence and poor mixing due to 1) the identification being imposed on the loadings of y_t and f_t and 2) f_t and Λ_m being sampled in different steps. That could be one of the reasons why the joint estimation gains were not significant in their paper compared to the sequential estimation. This is confirmed in the simulation study where neither the two-step PCA nor the one-step Bayesian approach recovers the unobserved factor produced through a simple FAVAR model. Therefore, this paper revisits a full Bayesian estimation (from here on referred to as BFAVAR) of the FAVAR model and proposes a precision-based algorithm which 1) uses the identification restrictions as suggested in Bai et al. (2016) and 2) represents the FAVAR as a dynamic factor model as in Chan and Jeliazkov (2009) for efficiency. BFAVAR draws factors and their loadings jointly, thus reducing the inefficiency for the Gibbs sampler and avoiding slow convergence and poor mixing. The simulation study shows that the BFAVAR can recover the factor from a simple FAVAR framework whereas both the estimation algorithms from BBE could not.

The remainder of the paper is structured as follows. Section 3.2 presents the representations of the FAVAR model which will be used for full conditionals in the Gibbs sampler. Section 3.3 discusses identification restrictions. Section 3.4 lays out the estimation procedure, which uses an efficient fitting algorithm based on MCMC simulation. Section 3.5 discusses the estimation of the marginal likelihood. Section 3.6 presents the simulation study comparing BFAVAR and BBE. Section 3.7 concludes this article.

3.2 Representations for Full Conditionals

The paper describes three representations of the FAVAR model which will be used to specify the full conditionals for unobservable factors and model parameters.

3.2.1 Representation 1

The first representation would be used to draw Γ, Λ, b_p, δ_1 and δ_2 from their full conditional distributions. By stacking Eq. 3.1 over time, we obtain the FAVAR model in the matrix form.

$$\begin{aligned} Z &= \tilde{X}_1\delta_1 + \tilde{y}\Gamma + \tilde{f}\Lambda + \varepsilon_1 \\ fy &= \tilde{X}_2\delta_2 + \sum_{p=1}^{P-1} \widetilde{fy}_p b_p + \varepsilon_2 \end{aligned} \tag{3.2}$$

where $b_p = vec(B_p')$,

$$\widetilde{fy}_p = \begin{pmatrix} I_{r_1+r_2} \otimes \begin{pmatrix} f'_{P-p} & y'_{P-p} \end{pmatrix} \\ \vdots \\ I_{r_1+r_2} \otimes \begin{pmatrix} f'_{T-p} & y'_{T-p} \end{pmatrix} \end{pmatrix}, \quad fy = \begin{pmatrix} f_P \\ y_P \\ \vdots \\ f_T \\ y_T \end{pmatrix}, \quad \varepsilon_2 = \begin{pmatrix} \varepsilon_{2P} \\ \vdots \\ \varepsilon_{2,T} \end{pmatrix},$$

$$\Gamma = \begin{pmatrix} \Gamma_1' \\ \vdots \\ \Gamma_n' \end{pmatrix}, \ \varepsilon_1 = \begin{pmatrix} \varepsilon_{1P} \\ \vdots \\ \varepsilon_{1,T} \end{pmatrix}, \ \Lambda = \begin{pmatrix} \Lambda_1' \\ \vdots \\ \Lambda_n' \end{pmatrix}, \ Z = \begin{pmatrix} Z_P \\ \vdots \\ Z_T \end{pmatrix},$$

$$\tilde{X}_1 = \begin{pmatrix} \tilde{X}_{1P} \\ \vdots \\ \tilde{X}_{1T} \end{pmatrix}, \ \tilde{X}_2 = \begin{pmatrix} \tilde{X}_{2P} \\ \vdots \\ \tilde{X}_{2T} \end{pmatrix}, \ \tilde{y} = \begin{pmatrix} I_n \otimes y_P' \\ \vdots \\ I_n \otimes y_T' \end{pmatrix}, \ \tilde{f} = \begin{pmatrix} I_n \otimes f_P' \\ \vdots \\ I_n \otimes f_T' \end{pmatrix}$$

3.2.2 Representation 2

For the second representation, the paper specifies the FAVAR model as a dynamic factor model (Chan and Jeliazkov (2009)) which will facilitate drawing factors from their full conditionals

$$\begin{aligned} W_t &= \tilde{X}_{3t}\delta_3 + \Gamma_w y_t + \Lambda_w f_t + \varepsilon_{3t} \\ f_t &= \tilde{X}_{4t}\delta_4 + \sum_{p=1}^{P-1} B_p^{fy} y_{t-p} + \sum_{p=1}^{P-1} B_p^{ff} f_{t-p} + e_t \end{aligned} \tag{3.3}$$

where in Eq. 3.3, the terms for $t = P, \ldots, T-1$ are defined as,

$$W_t = \begin{pmatrix} Z_t \\ y_{t+1} \end{pmatrix}, \; \tilde{X}_{3t} = \begin{pmatrix} X'_{1t} \otimes I_n & 0 \\ 0 & X'_{2t} \otimes I_{r_2} \end{pmatrix}, \; \Gamma_w = \begin{pmatrix} \Gamma & 0 & 0 \\ B_1^{yy} & \cdots & B_P^{yy} \end{pmatrix}, \; \delta_3 = \begin{pmatrix} \delta_1 \\ \delta_2^y \end{pmatrix},$$

$$\Lambda_w = \begin{pmatrix} \Lambda & 0 & 0 \\ B_1^{yf} & \cdots & B_P^{yf} \end{pmatrix} \text{ and } \varepsilon_{3t} = \begin{pmatrix} \varepsilon_{1t} \\ v_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega_{vv} \end{pmatrix} \right) = N(0, \Sigma_w)$$

and for $t = T$, the terms are defined as $W_T = Z_T$, $\tilde{X}_{3t} = X'_{1t} \otimes I_n$, $\delta_3 = \delta_1$, $\varepsilon_{3T} = \varepsilon_{1T}$, $\tilde{X}_{3T} = I_n \otimes X'_{1T}$, $\delta_3 = \delta_1$, $\Sigma_w = \Sigma$,

$$\Lambda_w = \begin{pmatrix} \Lambda & 0 & 0 \end{pmatrix} \text{ and } \Gamma_w = \begin{pmatrix} \Gamma & 0 & 0 \end{pmatrix}.$$

For the VAR portion of Eq. 3.3, $\tilde{X}_{4t} = X'_{2t} \otimes I_{r_1}$, $\delta_4 = \delta_2^f$ for $t = P, \ldots, T$. To facilitate the drawing of the factors from their full joint conditional distribution, the measurement equation of the FAVAR model in Eq. 3.3 is stacked over t and is represented as

$$W = \tilde{X}_3 \delta_3 + \tilde{\Gamma}_w y + \tilde{\Lambda}_w f + \varepsilon_3, \quad \varepsilon_3 \sim N(0, \tilde{\Sigma}_w) \tag{3.4}$$

where $\tilde{\Sigma}_w = \Sigma_w \otimes I_{T-P}$ with $\Sigma_w = \Sigma$ for $(T-P)^{th}$ diagonal block,

$$\tilde{\Gamma}_w = \begin{pmatrix} \Gamma & & & & & & & \\ B_1^{yy} & B_2^{yy} & \cdots & B_P^{yy} & & & & \\ & \Gamma & & & & & & \\ & B_1^{yy} & B_2^{yy} & \cdots & B_P^{yy} & & & \\ & & & \ddots & & \ddots & & \\ & & & & \Gamma & & & \\ & & & & B_1^{yy} & B_2^{yy} & \cdots & B_P^{yy} \\ & & & & & \Gamma & & \end{pmatrix}, \; \tilde{X}_3 = \begin{pmatrix} \tilde{X}_{3P} \\ \vdots \\ \tilde{X}_{3T} \end{pmatrix}$$

$$\tilde{\Lambda}_w = \begin{pmatrix} \Lambda & & & & & & & \\ B_1^{yf} & B_2^{yf} & \cdots & B_P^{yf} & & & & \\ & \Lambda & & & & & & \\ & B_1^{yf} & B_2^{yf} & \cdots & B_P^{yf} & & & \\ & & & \ddots & & \ddots & & \\ & & & & \Lambda & & & \\ & & & & B_1^{yf} & B_2^{yf} & \cdots & B_P^{yf} \\ & & & & & \Lambda & & \end{pmatrix}, \; \varepsilon_3 = \begin{pmatrix} \varepsilon_{3P} \\ \vdots \\ \varepsilon_{3T} \end{pmatrix},$$

$$W = \begin{pmatrix} W_P \\ \vdots \\ W_T \end{pmatrix}, \; y = \begin{pmatrix} y_P \\ \vdots \\ y_T \end{pmatrix} \text{ and } \; f = \begin{pmatrix} f_P \\ \vdots \\ f_T \end{pmatrix}$$

Now stacking the VAR portion of Eq. 3.3 as well

$$\begin{pmatrix} f_P \\ \vdots \\ f_T \end{pmatrix} = \begin{pmatrix} \tilde{X}_{4,P} \\ \vdots \\ \tilde{X}_{4T} \end{pmatrix} \delta_4 + \sum_{p=1}^{P} B_p^{ff} \otimes I_{T-P} \begin{pmatrix} f_{P-p} \\ \vdots \\ f_{T-p} \end{pmatrix} + \sum_{p=1}^{P} B_p^{fy} \otimes I_{T-P} \begin{pmatrix} y_{P-p} \\ \vdots \\ y_{T-p} \end{pmatrix} + \begin{pmatrix} e_P \\ \vdots \\ e_T \end{pmatrix}.$$

A finite prior distribution for the initial states of factor, $\{f_1, \ldots, f_{P-1}\}$, is necessary to have a proper distribution. We consider $f_p \sim N(0, \Omega_{ee})$ for all $p = 1, \ldots, P-1$. A simple change of variable will lead to a concise representation of the full conditional of f.

$$\begin{aligned} Ff &= \tilde{X}_4 \delta_4 + \sum_{p=1}^{P-1} (B_p^{fy} \otimes I_{T-P}) y_{-p} + e \\ f &= F^{-1} \tilde{X}_4 \delta_4 + F^{-1} \sum_{p=1}^{P-1} (B_p^{fy} \otimes I_{T-P}) y_{-p} + F^{-1} e \\ f &= \tilde{\mu}_f + F^{-1} e \end{aligned} \tag{3.5}$$

where $\tilde{\mu}_f = F^{-1}\tilde{X}_4\delta_4 + F^{-1}\sum_{p=1}^{P-1}(B_i^{ff} \otimes I_{T-P})y_{-p}$,

$$F = \begin{pmatrix} I_{r_2} & & & & & \\ & \ddots & & & & \\ & & I_{r_2} & & & \\ & & & & & \\ -B_p^{ff} & \cdots & -B_1^{ff} & I_{r_2} & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & -B_p^{ff} & \cdots & -B_1^{ff} & I_{r_2} \\ & & & & & \end{pmatrix}, \; \tilde{X}_4 = \begin{pmatrix} \tilde{X}_{4P} \\ \vdots \\ \tilde{X}_{4T} \end{pmatrix} \text{ and } y_{-p} = \begin{pmatrix} y_{P-p} \\ \vdots \\ y_{T-p} \end{pmatrix}$$

Thus, $f|\theta \sim N\Big(\tilde{\mu}_f, (F'S^{-1}F)^{-1}\Big)$. Finally, the second representation of the FAVAR model looks like

$$\begin{aligned} W &= \tilde{X}_3\delta_3 + \tilde{\Gamma}_w y + \tilde{\Lambda}_w f + \varepsilon_3 \\ f &= \tilde{\mu}_f + F^{-1}e \end{aligned}$$

3.2.3 Representation 3

The third representation is similar to the first representation but the VAR equation is been separated to facilitate specifying full conditional distributions for variance-covariance matrices.

$$
\begin{aligned}
Z &= \tilde{X}_1\delta_1 + \tilde{y}\Gamma + \tilde{f}\Lambda + \varepsilon_1 \\
y &= \tilde{X}_{2y}\delta_{2y} + \sum_{p=1}^{P-1} \widetilde{fyy}_p b_{yp} + v \\
f &= \tilde{X}_{2f}\delta_{2f} + \sum_{p=1}^{P-1} \widetilde{ffy}_p b_{fp} + e
\end{aligned}
\tag{3.6}
$$

where $b_{yp} = vec\Big(\begin{pmatrix} B_p^{yf} & B_p^{yy} \end{pmatrix}' \Big)$, $b_{fp} = vec\Big(\begin{pmatrix} B_p^{ff} & B_p^{fy} \end{pmatrix}' \Big)$

$$
\tilde{X}_{2y} = \begin{pmatrix} I_{r_2} \otimes X'_{2P} \\ \vdots \\ I_{r_2} \otimes X'_{2T} \end{pmatrix}, \quad \tilde{X}_{2f} = \begin{pmatrix} I_{r_1} \otimes X'_{2P} \\ \vdots \\ I_{r_1} \otimes X'_{2T} \end{pmatrix}
$$

$$
\widetilde{fyy}_p = \begin{pmatrix} I_{r_2} \otimes \begin{pmatrix} f'_{P-p} & y'_{P-p} \end{pmatrix} \\ \vdots \\ I_{r_2} \otimes \begin{pmatrix} f'_{T-p} & y'_{T-p} \end{pmatrix} \end{pmatrix} \text{ and } \widetilde{ffy}_p = \begin{pmatrix} I_{r_1} \otimes \begin{pmatrix} f'_{P-p} & y'_{P-p} \end{pmatrix} \\ \vdots \\ I_{r_1} \otimes \begin{pmatrix} f'_{T-p} & y'_{T-p} \end{pmatrix} \end{pmatrix}
$$

3.3 Identification

Equation 3.1 is not completely identifiable unless some restrictions are introduced. To be specific, the transition equation needs some restrictions for the identification of factors and their loadings.

For any $r_1 \times r_1$ matrix M_{11} and $r_1 \times r_2$ matrix M_{12}, one can rewrite the transition equation as

$$\begin{aligned} Z_t &= \tilde{X}_{1t}\delta_1 + \Gamma_m y_t + \Lambda_m f_t + \varepsilon_{1t} \\ &= \tilde{X}_{1t}\delta_1 + (\Gamma_m + \Lambda_m M_{12})y_t + (\Lambda_m M_{11})(M_{11}^{-1} f_t - M_{11}^{-1} M_{12} y_t) + \varepsilon_{1t} \\ &= \tilde{X}_{1t}\delta_1 + \Gamma_m^* y_t + \Lambda_m^* f_t^* + \varepsilon_{1t} \end{aligned}$$

where, $\Gamma_m^* = (\Gamma_m + \Lambda_m M_{12})$, $\Lambda_m^* = \Lambda_m M_{11}$ and $f_t^* = M_{11}^{-1} f_t - M_{11}^{-1} M_{12} y_t$. Equation 3.2 has two observably equivalent models. To uniquely identify Eq. 3.2, $r_1^2 + r_1 r_2$ restrictions have to be imposed since there are $r_1^2 + r_1 r_2$ free parameters in M_{11} and M_{12}. Bai et al. (2016) proved that $r_1^2 + r_1 r_2$ restrictions are necessary and sufficient to deal with the identification problem. They also proposed three sets of identification restrictions. The three sets of restrictions are

1. $\Omega_{ee} = I_{r_1}$, $\Omega_{ev} = 0$ and $\frac{1}{N}\Lambda_m \Omega_{\varepsilon\varepsilon}^{-1} \Lambda_m^{'} = Q$ where Q is a diagonal matrix with diagonal elements distinct and arranged in descending order and $\Omega_{\varepsilon\varepsilon} = E(\varepsilon_{1t}\varepsilon_{1t}')$.
2. $\Omega_{ee} = I_{r_1}$, $\Omega_{ev} = 0$ and the upper $r_1 \times r_1$ submatrix of Λ_m is lower triangular.
3. $\Omega_{ev} = 0$ and the upper $r_1 \times r_1$ submatrix of Λ is I_{r_1}.

Each of them impose $r_1^2 + r_1 r_2$ restrictions on the FAVAR model. All the set of restrictions imposes constraints on the structure of the variance of ε_{2t}. The first restriction also requires $\frac{1}{N}\Lambda_m E(\varepsilon_{1t}\varepsilon_{1t}')\Lambda_m$ to be diagonal which is generally used in the likelihood estimation (see Lawley and Maxwell (1971) and Bai et al. (2016)). The second set compels the upper $r_1 \times r_1$ submatrix of Λ_m to be a lower triangular matrix. Under this constraint, only the first unobservable factor affects the first variable, the first two unobservable factors affect the second variable, and so on. The third set restricts the upper $r_1 \times r_1$ matrix of Λ_m to be an identity matrix. In this case, the first unobservable factor affects only the first variable, the second unobservable factor affects only the second variable, and so on. Since more restrictions are imposed as compared to the second set of restrictions, no structure is required to be imposed on Ω_{ee}.

BBE in their seminal paper imposed different sets of restrictions than what was proposed in Bai et al. (2016). They restricted the upper $r_1 \times r_1$ submatrix of Λ to be I_{r_1} and the upper $r_2 \times r_2$ submatrix of Γ to be 0. They justified the constraints by assuming restrictions to the channels by which the y_t 's contemporaneously affect the Z_t. The estimation strategy proposed in this paper is generalizable to any identification strategy the researcher chooses.

3.4 Estimation

The paper considers the following priors for the estimation: $\Gamma \sim N(\Gamma_0, \Omega_{\Gamma 0})$, $\Lambda \sim N(\Lambda_0, \Omega_{\Lambda 0})$, $b_p \sim TN_{\{B_p \in D\}}(b_{p0}, \Omega_{b_{p0}})$, $\delta_1 \sim N(\delta_{10}, S_{10})$, $\delta_2 \sim N(\delta_{20}, S_{20})$, $\Sigma^{-1} \sim W(\sigma_0, \Sigma_0)$, $\Omega_{ee}^{-1} \sim W(\omega_{ee0}, \Omega_{ee0})$ and $\Omega_{vv}^{-1} \sim W(\omega_{vv0}, \Omega_{vv0})$. The issue with using the full conditional for unobserved factor loadings (Λ) for sampling is that the mixing takes a lot of time since factors and loadings are both unobserved. Chan and Jeliazkov (2009) suggested more efficient mixing schemes where factors and loadings are drawn in the same block using densities $p(f|\theta, W)$ and $p(\Lambda|\theta_{-\{f\}}, W)$. The density $p(\Lambda|\theta_{-\{f\}}, W)$ can be obtained as

$$\begin{aligned} p(\Lambda|\theta_{-\{f\}}, W) &= \frac{p(\Lambda|\theta, W)p(f|\theta_{-\{\Lambda_c\}}, W)}{p(f|\theta, W)} \\ p(\Lambda|\theta_{-\{f\}}, W) &\propto \frac{p(\Lambda|\theta, W)}{p(f|\theta, W)} \end{aligned}$$

where the full conditional densities $p(\Lambda|\theta, W)$ and $p(f|\theta, W)$ are known. Since $p(\Lambda|\theta_{-\{f\}}, W)$ does not have a standard form, we will use the MH algorithm for drawing. MCMC estimation can proceed through iterative sampling of the following steps.

FAVAR Model: MCMC Implementation

STEP 1. Sample $f, \Lambda|W, \theta$ in a single block to achieve efficiency.

(a) Sample $f|W,\theta \sim N(\tilde{f},\tilde{F})$ where,

$$\begin{aligned}\tilde{F} &= \Big(F'S^{-1}F + \tilde{\Lambda}'_w\tilde{\Sigma}_w^{-1}\tilde{\Lambda}_w\Big)^{-1} \\ \tilde{f} &= \tilde{F}\Big(F'S^{-1}\tilde{\mu}_f + \tilde{\Lambda}'_w\tilde{\Sigma}_w^{-1}(W - \tilde{X}_3\delta_3 - \tilde{\Gamma}_w y)\Big)\end{aligned}$$

(b) Sample $\Lambda|W,\theta_{\{-f\}}$ by MH algorithm with tailored proposal density $\Lambda \sim q(\hat{\Lambda},\Omega_\Lambda)$, a multivariate Student's t density with low degrees of freedom to ensure heavy tails where $\tilde{\Lambda}$ and $\tilde{L}$ are mode and inverse of the negative hessian at the mode of $p(\Lambda|\theta_{-\{f\}})$. Accept the proposed draw Λ^* with probability

$$\alpha_{MH}(\Lambda,\Lambda^*) = \min\left\{1, \frac{p(\Lambda^*|\theta_{-\{f\}},W)q(\Lambda|\hat{\Lambda},\Omega_\Lambda)}{p(\Lambda|\theta_{-\{f\}},W)q(\Lambda^*|\hat{\Lambda},\Omega_\Lambda)}\right\}$$

STEP 2. Sample $\Gamma|W,\theta \sim N(\hat{\Gamma},\Omega_\Gamma)$ where

$$\begin{aligned}\Omega_\Gamma &= (\Omega_{\Gamma 0}^{-1} + \tilde{y}'(I_{T-P}\otimes\Sigma)^{-1}\tilde{y})^{-1} \\ \hat{\Gamma} &= \Omega_\Gamma(\Omega_{\Gamma 0}^{-1}\Gamma_0 + \tilde{y}'(I_{T-P}\otimes\Sigma)^{-1}(Z - \tilde{X}_1\delta_1 - \tilde{f}\Lambda))\end{aligned}$$

STEP 3. Sample $b_p|W,\theta \sim N(\hat{b}_p,\Omega_{b_p})$ where

$$\begin{aligned}\Omega_{b_p} &= (\Omega_{b_{p0}}^{-1} + \widetilde{fy}'_p(I_{T-P}\otimes\Omega^{-1})\widetilde{fy}_p)^{-1} \\ \hat{b}_p &= \Omega_{b_p}(\Omega_{b_{p0}}^{-1}b_{p0} + \widetilde{fy}'_p(I_{T-P}\otimes\Omega^{-1})(fy - \tilde{X}_2\delta_2 - \sum_{j\neq p}\widetilde{fy}_j b_j))\end{aligned}$$

STEP 4. Sample $\delta_1|\theta_{\{-\delta_1\}} \sim N(\hat{\delta_1},\hat{S_1})$ where

$$\begin{aligned}\hat{S}_1 &= \Big(S_{10}^{-1} + \tilde{X}_1'(I_{T-P}\otimes\Sigma)\tilde{X}_1\Big)^{-1} \\ \hat{\delta}_1 &= \hat{S}_1\Big(S_{10}^{-1}\delta_{10} + \tilde{X}_1'(I_{T-P}\otimes\Sigma)(Z - \tilde{y}\Gamma - \tilde{f}\Lambda)\Big)\end{aligned}$$

STEP 5. Sample $\delta_2|\theta_{\{-\delta_2\}} \sim N(\hat{\delta}_2, \hat{S}_2)$ where

$$\begin{aligned}\hat{S}_2 &= \left(S_{20}^{-1} + \tilde{X}_2'(I_{T-P} \otimes \Omega)\tilde{X}_2\right)^{-1} \\ \hat{\delta}_2 &= \hat{S}_2\left(S_{20}^{-1}\delta_{20} + \tilde{X}_2'(I_{T-P} \otimes \Omega)(fy - \sum_{p=1}^{P} \widetilde{fy}_p b_p)\right)\end{aligned}$$

STEP 6. Sample $\Sigma^{-1}|W, \theta \sim W(\sigma_0 + T - P, (\Sigma_0^{-1} + \sum_{t=P}^{T} \varepsilon_{1t}\varepsilon_{1t}')^{-1})$.

STEP 7. Sample $\Omega_{ee}^{-1}|W, \theta \sim W(\omega_{ee0} + T - P, (\Omega_{ee0}^{-1} + \sum_{t=P}^{T} v_t v_t')^{-1})$.

STEP 8. Sample $\Omega_{vv}^{-1}|W, \theta \sim W(\omega_{vv0} + T - P, (\Omega_{vv0}^{-1} + \sum_{t=P}^{T} e_t e_t')^{-1})$.

3.5 Marginal Likelihood

The integrated likelihood for the FAVAR model takes the form

$$p(W|\theta) = \frac{p(W|f, \theta)p(f|\theta)}{\pi(f|W, \theta)}$$

where,

$$\begin{aligned}p(W|f, \theta) &= N(\tilde{X}_3\delta_3 + \tilde{\Gamma}_w y + \tilde{\Lambda}_w f, \tilde{\Sigma}_w) \\ p(f|\theta) &= N(\tilde{\mu}_f, (F'S^{-1}F)^{-1}) \\ p(f|W, \theta) &= N(\tilde{f}, \tilde{F})\end{aligned}$$

Using Chib (1995), the marginal likelihood for the FAVAR model can be estimated as

$$p(W) = \frac{p(W|\theta^*)\pi(\theta^*)}{\pi(\theta^*|W)}$$

where θ^* is the parameter values evaluated as the mode of the posterior distribution. Substituting integrated likelihood value in marginal likelihood gives the following form

$$p(W) = \frac{p(W|f,\theta^*)p(f|\theta^*)\pi(\theta^*)}{\pi(f|W,\theta^*)\pi(\theta^*|W)}$$

where $\pi(\theta)$ is the priors distribution and $\pi(\theta|W)$ is the posterior distribution marginalized for factors which can be estimated as

$$\pi(\theta|W) = \frac{1}{G}\sum_{g=1}^{G}\pi(\theta|W, f^{\{g\}})$$

where $\pi(\theta|W, f)$ is the posterior distribution of the model parameters and $\{f^{\{g\}}\}$ represent g^{th} draw from the main MCMC run.

3.6 Simulation Study

This section compares the estimation performance of BFAVAR introduced in this paper with the one-step Bayesian and two-step PCA estimation techniques from BBE. The data is artificially generated using a simple FAVAR model with one factor, one lag, no covariates and no intercept. The observed factor y_t is 3×1 dimensional whereas Z_t is 20×1 dimensional where $t = 1, \ldots, 100$. The data is simulated from the following FAVAR model.

$$Z_t = \Gamma_m y_t + \Lambda_m f_t + \varepsilon_{1t}$$

$$\begin{pmatrix} y_t \\ f_t \end{pmatrix} = B_1 \begin{pmatrix} y_{t-1} \\ f_{t-1} \end{pmatrix} + \varepsilon_{2t},$$

where $\Sigma = I_{20} * 0.1$,

$$\Lambda_m = \begin{pmatrix} -0.58 & 0.86 & -0.86 \\ 0.28 & 0.34 & -0.43 \\ -0.33 & 0.40 & -0.17 \\ 0.90 & 0.69 & 0.88 \\ 0.63 & 0.41 & 0.95 \\ 0.48 & 0.71 & -0.91 \\ -0.55 & -0.18 & 0.17 \\ -0.90 & 0.35 & -0.93 \\ -0.26 & -0.74 & 0.13 \\ 0.48 & 0.47 & -0.80 \end{pmatrix}, \quad \Gamma_m = \begin{pmatrix} 1.00 \\ 0.63 \\ -0.30 \\ -0.72 \\ -0.83 \\ -0.30 \\ 0.98 \\ 0.065 \\ -0.32 \\ -0.37 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} 0.3 & -0.2 & 0.1 & 0.2 \\ 0.3 & 0.4 & 0.3 & -0.3 \\ -0.3 & 0.3 & 0.2 & 0.1 \\ 0.2 & -0.2 & 0.2 & 0.3 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1.0 & 0.0 & 0.6 & 0.0 \\ 0.0 & 1.0 & 0.3 & 0.0 \\ 0.6 & 0.3 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

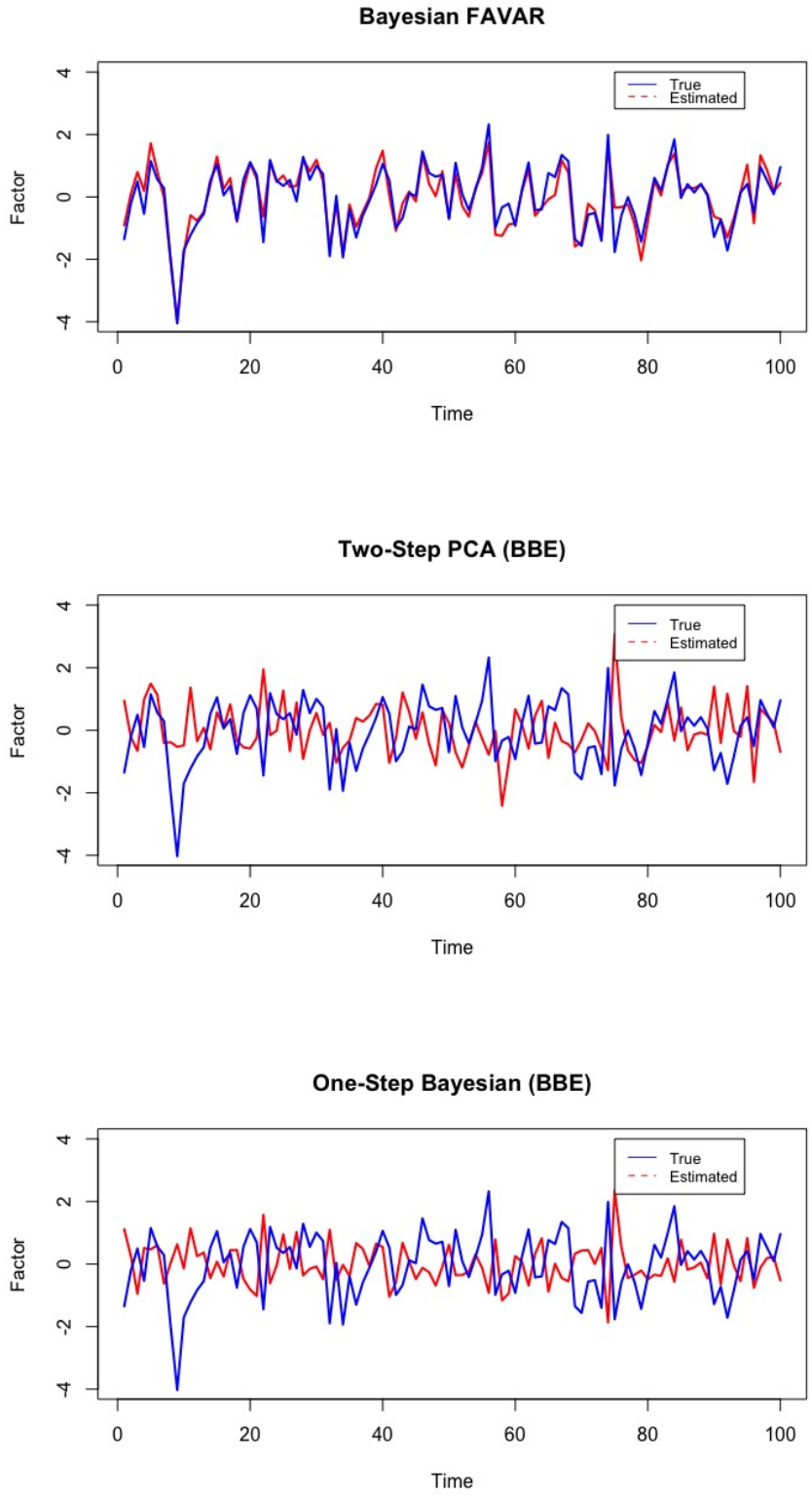


Figure 3.1: Estimated Factors from different Bayesian Estimation Approaches

The study adopts a variation of restriction (2) which means $\Omega_{ee} = \sigma^2 I_{r_1}$, $\Omega_{ev} = 0$ and the upper $r_1 \times r_1$ submatrix of Λ_m to be lower triangular with diagonal elements equal to 1. The FAVAR model, generated via the DGP, is estimated using BFAVAR and BBE estimation techniques. Figure 3.1 has three subfigures, each presenting the estimated factor via the concerned estimation technique superimposed over the true factor. The Bayesian FAVAR successfully recovers the unobserved factor, whereas the one-step and two-step BBE techniques fails to do so. This simulation study confirms the concerns raised in Sec. 3.1 regarding BBE estimation techniques and proposes further enquiry to identify the source of bias.

3.7 Conclusion

This paper examines the Bayesian estimation of the FAVAR model and proposes a precision-based Bayesian algorithm. Efficient MCMC sampling is discussed in the context of estimating factors where the prior distribution is derived from the VAR portion of the model Implementation of BFAVAR in other simulation settings and real life applications is an interesting area for future research.

Bibliography

Aastveit, K. A., Mitchell, J., Ravazzolo, F., and Van Dijk, H. K. (2018). The evolution of forecast density combinations in economics. Technical report, Tinbergen Institute Discussion Paper.

Amir-Ahmadi, P. and Uhlig, H. (2009). Measuring the dynamic effects of monetary policy shocks: A bayesian favar approach with sign restriction.

Bacharach, J. (1974). Bayesian dialogues. *Unpublished manuscript, Christ Church College, Oxford University*.

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of favar models. *Journal of Business & Economic Statistics*, 34(4):620–641.

Balke, N. S. (2000). Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, 82(2):344–349.

Ball, L., Mankiw, N. G., Romer, D., Akerlof, G. A., Rose, A., Yellen, J., and Sims, C. A. (1988). The new keynesian economics and the output-inflation trade-off. *Brookings papers on economic activity*, 1988(1):1–82.

Bassetti, F., Casarin, R., and Ravazzolo, F. (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, 113(522):675–685.

Baştürk, N., Borowska, A., Grassi, S., Hoogerheide, L., and van Dijk, H. K. (2019). Forecast density combinations of dynamic models and data driven portfolio strategies. *Journal of Econometrics*, 210(1):170–186.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Bernardo, J. M. and Smith, A. F. (2000). *Bayesian theory*, volume 405. John Wiley & Sons.

Bianchi, F., Mumtaz, H., and Surico, P. (2009). Dynamics of the term structure of uk interest rates.

Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2):213–232.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *American economic review*, 99(1):350–384.

Busetti, F. (2017). Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics*, 79(4):495–512.

Caggiano, G., Castelnuovo, E., and Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in us recessions. *Journal of Monetary Economics*, 67:78–92.

Canova, F. (1993). Modelling and forecasting exchange rates with a bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control*, 17(1-2):233–261.

Capistrán, C. and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, 41(2-3):365–396.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *the Quarterly Journal of economics*, 118(1):269–298.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association*, 90(432):1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75(1):79–97.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183–206.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American statistical association*, 96(453):270–281.

Chib, S. and Jeliazkov, I. (2005). Accept–reject metropolis–hastings sampling and marginal likelihood estimation. *Statistica Neerlandica*, 59(1):30–44.

Chib, S. and Jeliazkov, I. (2006). Inference in semiparametric dynamic models for binary longitudinal data. *Journal of the American Statistical Association*, 101(474):685–700.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

Clements, M. P. (2002). An evaluation of the survey of professional forecasters probability distributions of expected inflation and output growth. *Manuscript, Department of Economics, University of Warwick*.

Clements, M. P., Rich, R. W., and Tracy, J. S. (2023). Surveys of professionals. In *Handbook of Economic Expectations*, pages 71–106. Elsevier.

Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical science*, 19(1):81–94.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics*, 8(2):262–302.

Coibion, O., Gorodnichenko, Y., and Kamdar, R. (2018). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature*, 56(4):1447–91.

Conflitti, C., De Mol, C., and Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4):1096–1103.

Croushore, D., Stark, T., et al. (2019). Fifty years of the survey of professional forecasters. *Economic Insights*, 4(4):1–11.

Croushore, D. D. (1993). Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia*, 6:3.

Dahl, C. M. and González-Rivera, G. (2003-1). Testing for neglected nonlinearity in regression models based on the theory of random fields. *Journal of Econometrics*, 114(1):141–164.

Dahl, C. M. and Gonzalez-Rivera, G. (2003-2). Identifying nonlinear components by random fields in the us gnp growth. implications for the shape of the business cycle. *Studies in Nonlinear Dynamics & Econometrics*, 7(1).

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, pages 65–81.

Degroot, M. H. and Mortera, J. (1991). Optimal linear opinion pools. *Management Science*, 37(5):546–558.

Del Negro, M., Hasegawa, R. B., and Schorfheide, F. (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics*, 192(2):391–405.

Diebold, F. X., Tay, A., and Wallis, K. (1997). Evaluating density forecasts of inflation: the survey of professional forecasters.

Dueker, M. (2005). Dynamic forecasts of qualitative variables: a qual var model of us recessions. *Journal of Business & Economic Statistics*, 23(1):96–104.

Eickmeier, S., Gambacorta, L., and Hofmann, B. (2014). Understanding global liquidity. *European Economic Review*, 68:1–18.

Elliott, G., Ghanem, D., and Krüger, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Review of Economics and Statistics*, 98(4):742–755.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Fildes, R. and Ord, K. (2002). Forecasting competitions–their role in improving forecasting practice and research. *A companion to economic forecasting*, pages 322–353.

Forni, M. and Gambetti, L. (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203–216.

Friedman, M. (1968). The role of monetary policy the american economic review. *New york*, 58.

Garratt, A., Henckel, T., and Vahey, S. P. (2023). Empirically-transformed linear opinion pools. *International Journal of Forecasting*, 39(2):736–753.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

Geweke, J. and Amisano, G. (2012). Prediction with misspecified models. *American Economic Review*, 102(3):482–86.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Gneiting, T. and Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7:1747–1782.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68(1):189–196.

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*, pages 357–384.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of policy Modeling*, 14(4):517–533.

Härdle, W., Mammen, E., and Müller, M. (1998a). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, 93(444):1461–1474.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*, 81(1):223–242.

Härdle, W., Tsybakov, A., and Yang, L. (1998b). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, 68(2):221–245.

Hastie, T. and Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, pages 1005–1016.

Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *Statistical science*, 14(4):382–417.

Huang, B.-N., Hwang, M.-J., and Peng, H.-P. (2005). The asymmetry of the impact of oil price shocks on economic activities: an application of the multivariate threshold model. *Energy Economics*, 27(3):455–476.

Jeliazkov, I. (2013). Nonparametric vector autoregressions: Specification, estimation, and inference. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Jeliazkov, I. and Lee, E. H. (2010). Mcmc perspectives on simulated likelihood estimation. In *Maximum simulated likelihood methods and applications*. Emerald Group Publishing Limited.

Kalli, M. and Griffin, J. E. (2018). Bayesian nonparametric vector autoregressive models. *Journal of econometrics*, 203(2):267–282.

Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review*, pages 714–735.

Kenny, G., Kostka, T., and Masera, F. (2015). Density characteristics and density forecast performance: a panel analysis. *Empirical Economics*, 48:1203–1231.

Kim, C.-J., Morley, J., and Piger, J. (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics*, 20(2):291–309.

Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80(2):188–201.

Koop, G. and Poirier, D. J. (2004). Bayesian variants of some classical semiparametric regression techniques. *Journal of Econometrics*, 123(2):259–282.

Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370.

Lawley, D. N. and Maxwell, A. E. (1971). Factor analysis as a statistical method.

Lindley, D. (1971). Bayesian statistics: A review (siam).

Long Jr, J. B. and Plosser, C. I. (1983). Real business cycles. *Journal of political Economy*, 91(1):39–69.

Ludvigson, S. C. and Ng, S. (2009). A factor analysis of bond risk premia. Technical report, National Bureau of Economic Research.

McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110.

McAlinn, K. and West, M. (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics*, 210(1):155–169.

Mitchell, J. and Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the klic with an application to the bank of england and niesr 'fan'charts of inflation. *Oxford bulletin of economics and statistics*, 67:995–1033.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146(1):26–43.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29(1):46–75.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pages 315–335.

Opschoor, A., Van Dijk, D., and van der Wel, M. (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics*, 32(7):1298–1313.

Phelps, E. S. (1967). Phillips curves, expectations of inflation and optimal unemployment over time. *Economica*, pages 254–281.

Poirier, D. J. (1998). Revising beliefs in nonidentified models. *Econometric theory*, 14(4):483–509.

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

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207.

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.

Sims, C. A. et al. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, 10(Win):2–16.

Sims, C. A. and Zha, T. (2006). Were there regime switches in us monetary policy? *American Economic Review*, 96(1):54–81.

Smets, F., Warne, A., and Wouters, R. (2014). Professional forecasters and real-time forecasting with a dsge model. *International Journal of Forecasting*, 30(4):981–995.

Steel, M. F. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, 58(3):644–719.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbooks Online*.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of monetary economics*, 44(2):293–335.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics*, pages 1339–1342.

Swanson, N. R. and White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International journal of Forecasting*, 13(4):439–461.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.

Tong, H. (1978). On a threshold model.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3):245–268.

Tsai, H. and Tsay, R. S. (2010). Constrained factor models. *Journal of the American Statistical Association*, 105(492):1593–1605.

Van Robays, I. (2016). Macroeconomic uncertainty and oil price volatility. *Oxford Bulletin of Economics and Statistics*, 78(5):671–693.

Vieu, P. (1995). Order choice in nonlinear autoregressive models. *Statistics: A Journal of Theoretical and Applied Statistics*, 26(4):307–328.

Vigfusson, R. (1997). Switching between chartists and fundamentalists: a markov regime-switching approach. *International Journal of Finance & Economics*, 2(4):291–305.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):364–372.

Wallis, K. F. (2005). Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, 67:983–994.

Wang, H., Zhang, X., and Zou, G. (2009). Frequentist model averaging estimation: a review. *Journal of Systems Science and Complexity*, 22(4):732–748.

Winkler, R. L., Munoz, J., Cervera, J. L., Bernardo, J. M., Blattenberger, G., Kadane, J. B., Lindley, D. V., Murphy, A. H., Oliver, R. M., and Ríos-Insua, D. (1996). Scoring rules and the evaluation of probabilities. *Test*, 5(1):1–60.

Appendix A

Chapter 1

A.1 Back-fitting MCMC Algorithm

This subsection presents the MCMC algorithm used in the estimation of BOP. Let w_o be the weight vector drawn in the previous iteration and w_n be the weight vector drawn in the current iteration. The steps are as follows.

STEP 1. Draw w_n from a proposal density, be it Dirichlet, Normal distribution with logistic transformation (discussed in Section 1.3) or truncated normal (defined on the interval $[0, 1]$), where the proposal is centred at w_o. Normalize w_n so that the sum is 1 in case needed, and choose the variance so that the whole space can be explored.

STEP 2. Generate $u_2 \sim uniform(0, 1)$

STEP 3. If $u_2 \leq min\Big(\frac{p(Y_T|w_n)}{p(Y_T|w_o)}, 1\Big)$, return w_n, else return w_o and store the value of conditional density evaluated at w_o. Since uniform Dirichlet distribution is considered as prior, it disappears from the formula.

STEP 4. Repeat the above three steps M times (call it iteration cycle 1) and name the weights as w_0^* with the highest conditional density value.

STEP 5. Repeat the above 4 steps N times (call it iteration cycle 2) with $w_0 = w_0^*$ in each iteration. Stop once the value of conditional density has converged and use w_0^* stored in the N^{th} iteration as $\bar{w_T}$.

The value N in iteration cycle 2 can be decided based on how much the maximum conditional density value changes after every M iteration in iteration cycle 1. Similarly, the number of iterations M in iteration cycle 1 is decided based on the trade-off between exploring the solution space and computational time. There is a possibility that $\bar{w_T}$ is not a global maximum. The paper suggests using the algorithm multiple times from different initial conditions to verify.

A.2 Asymptotic Properties

Under the M-closed case, when the true model (let's say D) is part of the set of available models, the opinion pool degenerates to the true model since all the weight is allotted to it (Geweke and Amisano (2011)). This situation rarely arrives in real life, and D is generally unknown to the forecaster and the decision maker. The weights become relevant under the M-Open case when D is not part of the set of available models. In that case, the true weights (let's say $w^0 = \{w_1^0, w_2^0, \ldots, w_K^0\}$) can be interpreted as the ones which give the minimum Kullback-Leibler divergence from D to the opinion pool. Gneiting and Raftery (2007) showed that the opinion pool optimized based on log predictive score minimizes the Kullback–Leibler directed distance from the data generating process to the prediction model. For K prediction models, the log prediction score for an opinion pool for $w_T = \{w_{1,T}, w_{2,T}, \ldots, w_{K,T}\}$ where $w_{k,T} \geq 0 \quad \forall \quad k = 1, 2, \ldots, K$ and $\sum_{k=1}^{K} w_{k,T} = 1$ for a given period t will look like

$$\begin{aligned} l(w_T|Y_T) &= \sum_{t=1}^{T} log\Big(\sum_{k=1}^{K} w_{k,T}\, p(y_t|Y_{t-1}, M_k)\Big) \\ &= \sum_{t=1}^{T} l(w_T|Y_t) \end{aligned} \tag{B.1}$$

One of the advantages of the log prediction score is that it is closely related to the likelihood function, which can be seen in the relation $l(w_T|Y_T) = log(p(Y_T|w_T))$. Geweke and Amisano (2011) showed that the weights obtained from optimizing $l(w_T|Y_T)$ asymptotically minimizes the Kullback-Leibler distance from the true model D.

$$w_T^* = arg\ max_w\ l(w_T|Y_T) \overset{a.s.}{\rightarrow} arg\ max_w\ l(w|Y) = w^0 \tag{B.2}$$

where, $\frac{1}{T}\sum_{t=1}^{T} l(w_T|Y_t) = \bar{l}(w_T|Y_T) \overset{a.s.}{\rightarrow} l(w|Y)$. Using this result, the posterior density of weights can be rewritten as

$$\begin{aligned} p(w_T|Y_T) &\propto p(Y_T|w_T)p(w_T) \\ &\propto exp\{log(p(Y_T|w_T))\}p(w_T) \\ &\propto exp\{\sum_{t=1}^{T} l(w_T|Y_t)\}p(w_T) \\ &\propto exp\{T\bar{l}(w_T|Y_T)\}p(w_T) \end{aligned} \tag{B.3}$$

As T increases, the exponential term dominates, and the effect of the prior, which does not depend on T, becomes relatively smaller. To analyse the posterior density further, let's take a second-order Taylor series approximation of $l(w_T|Y_T)$ around w_T^*

$$\begin{aligned} l(w_T|Y_T) &\approx l(w_T^*|Y_T) - \frac{T}{2}(w_T - w_T^*)^2(-\bar{l}''(w_T^*|Y_T)) \\ &\approx l(w_T^*|Y_T) - \frac{T}{2v}(w_T - w_T^*)^2 \end{aligned} \tag{B.4}$$

where $\bar{l}''(w_T^*|Y_T) = \frac{1}{T}\sum_{t=1}^{T} l''(w_T^*|Y_t)$ and $v = [\bar{l}''(w_T^*|Y_T)]^{-1}$. The term with first-order derivative disappears as $l(w_T|Y_T)$ is maximized at $w_T = w_T^*$. The posterior density can be approximated as

$$p(w_T|Y_T) \quad \propto \quad exp\{-\frac{T}{2v}(w_T - w_T^*)^2\}p(w_T) \tag{B.5}$$

The first term is in the form of a normal distribution with mean w_T^* and variance $\frac{v}{T}$. In summary, the role of the prior density becomes relatively small in determining the posterior density when T is large. The posterior density converges to a degenerate density at w^0 as $T \longrightarrow \infty$ then $\frac{v}{T} \longrightarrow 0$ and $w_T^* \longrightarrow w^0$, and the posterior density is approximately normally distributed with mean w_T^*.