A quasi-Bayesian local likelihood approach to time varying parameter VAR models[*]

Katerina Petrova[†]

April 23, 2018

Abstract

The paper establishes a quasi-Bayesian local likelihood (QBLL) estimation methodology for a multivariate model with time varying parameters. The asymptotic validity of the resulting quasi-posterior distributions of the drifting parameters is proven in general and, in the special case of a Gaussian VAR model, a closed form time varying Normal-Wishart expression for the quasi-posterior distribution of the parameters is provided. In addition, this paper develops several Gibbs algorithms, which can sample from a VAR model with a mixture of time varying and time invariant parameters. The proposed estimators differ from existing state space approaches to VAR models in that they estimate parameter time variation nonparametrically, without imposing parametric stochastic processes on the parameters. The QBLL estimators are robust to misspecification of the state equation and exhibit good finite sample performance, even when compared to correctly specified parametric state space models, as illustrated by a Monte Carlo exercise. In addition, we demonstrate that the QBLL approach provides a remedy to the 'curse of dimensionality' by accommodating large dimensional VAR systems and delivers improvements in the out-of-sample forecasts of key macroeconomic variables. Finally, the paper makes an empirical contribution to the literature on changing inflation dynamics in the U.S., presenting evidence of a fall in inflation persistence and volatility during the Great Moderation period.

JEL codes: C11, E27, E52

Keywords: time varying parameters, changing volatility, monetary policy, Bayesian methods

---

[†]University of St Andrews, Tel: +44 (0)1334 461956, Email: Katerina.Petrova@st-andrews.ac.uk

1   Introduction

Time varying parameter models have recently received a lot of attention in macroeconometric literature, due to their ability to accommodate structural changes or breaks in the relationships between key macroeconomic variables. As the time span of available data increases, it is implausible that relationships between economic variables remain fixed; consequently, the issue of accommodating parameter time variation in econometric models, such as vector autoregressions (VARs), has assumed an increasingly prominent role both in theoretical and empirical macroeconometric research.

The standard approach to the estimation of time varying parameter (TVP) VAR models employs state space methods. In a seminal paper, Cogley and Sargent (2002) study the changing dynamics of macroeconomic variables in the U.S. using a TVP VAR with autoregressive coefficients modelled as random walk processes. In a remark, Sims (2001) and Stock (2001) point out the possibility of exaggerated parameter drift uncovered by Cogley and Sargent (2002) as a result of their time invariant covariance matrix assumption. In subsequent work, Primiceri (2005) and Cogley and Sargent (2005) address this issue by accommodating drifting volatility in VAR models, the former by employing a procedure suggested by Kim, Shephard and Chib (1998) and the latter by utilising a Metropolis within Gibbs algorithm with a technique from Jacquier, Polson and Rossi (1994). Both approaches require a parametric specification of the stochastic process generating the volatility parameters. More recently, Cogley, Primiceri and Sargent (2010) propose a VAR model which, in addition to drifts in the parameters and volatilities, also features time varying volatility in the state equations of the autoregressive parameters.

An alternative to the parametric state space approach is presented in Giraitis, Kapetanios and Yates (2014, 2018), who propose a nonparametric method for the estimation of the coefficient and variance processes in a time varying linear regression setting and establish the theoretical properties of their kernel-type estimator. Their method is extended to a general local likelihood framework that accommodates consistent and asymptotically normal extremum estimation by Giraitis, Kapetanios, Wetherilt and Zikes (2016).

Neither state space models nor the frequentist estimators of Giraitis et al. (2014, 2016, 2018) can accommodate large dimensional systems in the presence of time variation in the parameters. This problem is particularly relevant when considering VAR models which are richly parameterised even for a small number of endogenous variables and lags and, as a consequence, run the risk of overfitting the data. This paper attempts to address this issue by combining the existing nonpara-

metric approach with Bayesian methods and establishing a quasi-Bayesian local likelihood (QBLL) estimation methodology for a general multivariate model with time varying parameters. The QBLL approach augments the frequentist estimators of Giraitis et al. (2014, 2016, 2018) in order to provide a Bayesian treatment for the drifting parameters. We prove that the resulting quasi-posterior distributions are asymptotically valid for inference and confidence interval construction. In addition, we derive in closed form a Normal-Wishart expression for the quasi-posterior density in the special case of a Gaussian VAR model.

Another key issue addressed in this paper is the development of an econometric procedure that can estimate mixtures of time invariant and time varying parameters, by developing several Gibbs algorithms based on the analytic expressions obtained for the conditional quasi-posterior densities. It is the Bayesian treatment of this paper that facilitates the construction of such algorithms and, to our best knowledge, this is the first procedure that can accommodate such mixtures without imposing parametric assumptions on the parameter processes.

The proposed QBLL method has several advantages over the widely used state space approaches to drifting parameters. First, the parameter time variation enters nonparametrically, ensuring consistent estimation in a wide class of parameter processes, and alleviating the risk of invalid inference due to misspecification of the state equation. This point is illustrated further in the Monte Carlo exercise in Section 4. Second, unlike standard state space models, the proposed method does not suffer from dimensionality issues. The combination of our closed form quasi-posterior expressions with Minnesota-type priors, specified directly on the drifting parameters, allows the number of variables in the VAR to be very large while also facilitating the parameter drift. To illustrate this point, the QBLL approach can handle estimation of an 87 variable TVP VAR model (see Section 5.2), whereas state space methods are limited to models with at most 4 variables. The availability of analytic expressions for the quasi-posterior density in the Gaussian VAR case alleviates the computational burden of Markov chain Monte Carlo (MCMC) algorithms used for the estimation of TVP VARs by state space methods, making the proposed quasi-Bayesian procedure computationally efficient. Another advantage of our novel approach is the inverted-Wishart property of the time varying covariance matrix, which ensures symmetric positive definiteness of the resulting estimators without the need for further restrictions (e.g. triangularisation) commonly imposed by the state space literature; the absence of such ensuring that the ordering of the variables in the VAR system does not affect the estimation of the reduced form covariance matrix.

We apply our novel QBLL approach to study the changing dynamics in a VAR model with key

U.S. macroeconomic variables. We find considerable time variation in the series for the natural rate of unemployment and core inflation, as well as a significant decline in inflation persistence after the beginning of the Great Moderation period. These results confirm evidence presented in Cogley and Sargent (2002, 2005) on the time varying dynamics of inflation. We also uncover a fall in the volatilities of the series after the end of the oil crises, which is in line with evidence reported in Primiceri (2005) and Sims and Zha (2006).

In order to access the forecasting record of the QBLL estimators, we design an out-of-sample forecasting exercise. We find that allowing for time varying parameters can improve both point and density forecasts of Bayesian VAR models. Increasing the dimension of the VAR along with parameter time variation, made feasible by the use of the QBLL methodology established in this paper, can enhance forecast performance further, delivering unbiased forecasts and uniform probability integral transformations for most variables and horizons. Moreover, by accommodating mixtures of time varying and time invariant parameters, the algorithms developed in this paper allow us to 'switch off' the variation in the autoregressive or covariance matrices of the VAR respectively and investigate further the source of forecast improvements. To this end, we find that allowing for variation in the autoregressive parameters eliminates systematic forecast errors while drifts in the volatility improve density forecasts.

The rest of the paper is organised as follows. Section 2 establishes the QBLL methodology and proves its asymptotic validity. In Section 3.1, we derive closed form quasi-posterior expressions for the special case of a Gaussian VAR model; these expressions provide the building block of a number of Gibbs algorithms developed in Section 3.3. In Section 4, we study the finite sample properties of the QBLL estimators and compare them to existing methods in a Monte Carlo experiment. Section 5 contains our empirical contribution to the literature on changing dynamics of inflation in the U.S. and the forecasting exercise. Section 6 provides a concluding discussion and the Supplementary Appendix contains all proofs of propositions and some additional algorithms and empirical results.

2 The QBLL approach

This section establishes the QBLL methodology for inference in the presence of time varying parameters. Bayesian analysis is conducted by, first, employing a modification to the kernel weights used for Giraitis et al. (2014, 2016, 2018) frequentist estimators and, second, augmenting the resulting modified objective function by a prior density and deriving a posterior density for the purpose of

estimation and confidence interval construction. The asymptotic validity of the posterior density arising from the QBLL approach is formally established by Proposition 1 below.

Let y_t be an observed time series with log probability density $l_t(y_t|y^{t-1}, \theta_t)$, conditional on history $y^{t-1} = \{y_1, ..., y_{t-1}\}$, and indexed by a time varying finite-dimensional vector of parameters θ_t, satisfying one of the conditions (i)-(ii) presented below.

(i) θ_t is a deterministic process $\theta_t = \theta(t/T)$, where $\theta(.)$ is a piecewise differentiable function.

(ii) θ_t is a stochastic process satisfying: $\sup_{j:|j-t|\leq h} ||\theta_t - \theta_j||^2 = O_p(h/t)$ for $1 \leq h \leq t$ as $t \to \infty$.

Figure 1: Examples of parameter processes satisfying conditions (i) or (ii)
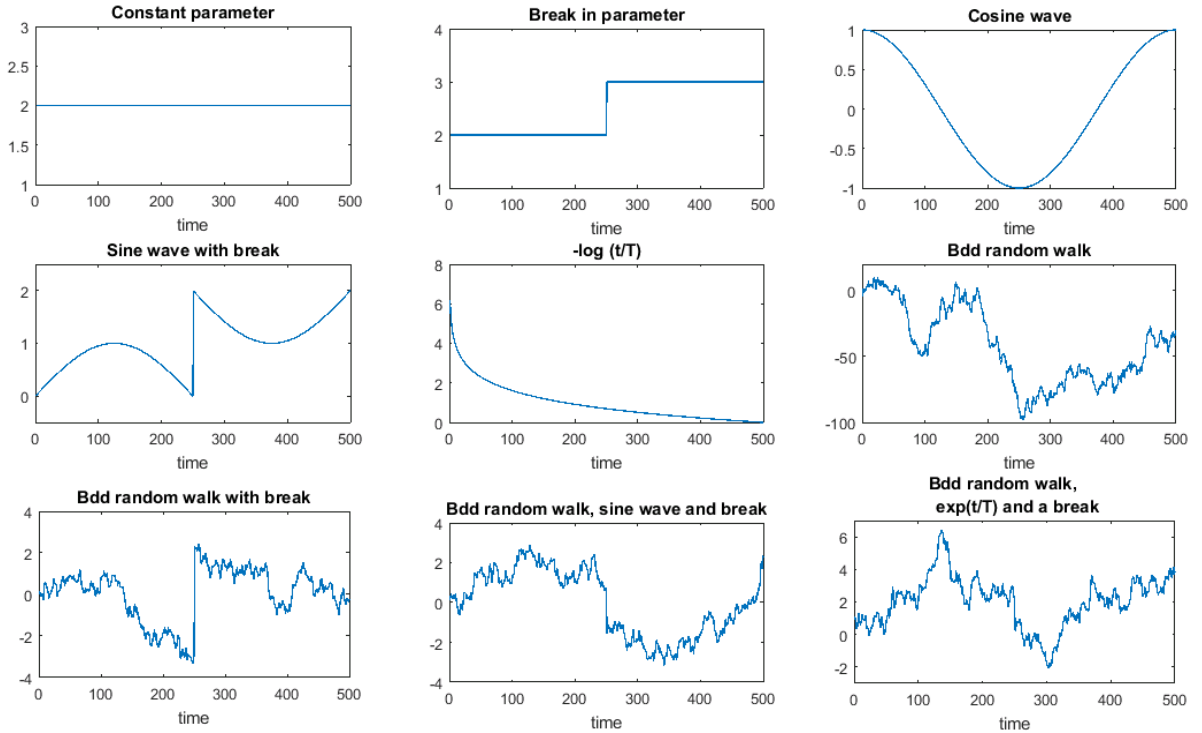
Both (i) and (ii) imply that the sequence of parameters drifts slowly with time, a property that is sufficient for consistent estimation of θ_t. Condition (i) is a standard assumption made in the work of Dahlhaus on locally stationary processes (e.g. Dahlhaus (2000) or Dahlhaus and Polonik (2006)) and later used by Giraitis et al. (2014, 2016, 2018). It requires that the parameter process is a piecewise smooth deterministic function, thus allowing for breaks in the parameters. Condition (ii) is a generalisation of (i) to include stochastic parameter processes exhibiting a degree of persistence necessary for consistent estimation of stochastically driven time variation. Condition (ii) includes the bounded random walk process used in the Monte Carlo exercise in Section 4, as well as some fractionally integrated processes. In addition, the parameters can feature any combination of

deterministic trends and/or breaks satisfying (i) and persistent stochastic components satisfying (ii). In Figure 1 we provide several illustrations of parameter processes that conform to conditions (i) and (ii). Under the parameter time variation framework of (i) and (ii), an extremum estimator $\hat{\theta}_{Tj} = \arg\max_\theta \ell_{Tj}(\theta_j)$ for θ_j is derived by maximising an objective function given by

$$\ell_{Tj}(\theta_j) := \sum_{t=1}^{T} w_{jt} l_t \left(y_t | y^{t-1}, \theta_j\right) \quad j \in \{1, ..., T\} \tag{1}$$

where $l_t\left(y_t | y^{t-1}, \theta_j\right)$ is the conditional log-density[1] for observation t and the weights w_{jt} are computed using a kernel function and normalised to sum to one:

$$w_{jt} = \tilde{w}_{jt} / \sum_{t=1}^{T} \tilde{w}_{jt}, \quad \tilde{w}_{jt} = \mathcal{K}\left(\frac{j-t}{H}\right) \quad \text{for} \quad j, t \in \{1, ..., T\}. \tag{2}$$

The kernel function $\mathcal{K}$ is assumed to be non-negative, continuous and bounded function with domain $\mathbb{R}$. The bandwidth parameter H satisfies $H \to \infty$ and $H = o(T/\log T)$. For example, the widely used Normal kernel weights are given by $\tilde{w}_{jt} = (1/\sqrt{2\pi}) \exp((-1/2)((j-t)/H)^2)$ for $t, j = 1, .., T$, while the rolling window procedure results as a special case of the choice of flat kernel weights: $w_{jt} = \mathbb{I}(|t-j| \leq H)$. For further discussion of the advantages of exponential kernels over the flat kernel for introducing time variation, refer to Giraitis et al. (2014) and Giraitis, Kapetanios and Price (2013). In this setup, Giraitis et al. (2016) show that, under regularity conditions, $\hat{\theta}_j$ is an $H^{1/2} + (T/H)^{1/2}$- consistent estimator[2] of θ_j for all $j = [\tau T]$ with some fixed $\tau \in (0, 1)$. Furthermore, defining

$$\widehat{\Sigma}_{Tj} := \frac{1}{\varkappa_{Tj}} \left(-\frac{\partial^2 \ell_{Tj}(\hat{\theta}_j)}{\partial\theta\partial\theta'}\right)^{-1}, \qquad \varkappa_{Tj} := \left(\sum_{t=1}^{T} w_{jt}^2\right)^{-1}, \tag{3}$$

a.s. positive definiteness of $\widehat{\Sigma}_{Tj}$ and the bandwidth rate $H = o(T^{1/2})$ are sufficient for asymptotic normality of $\hat{\theta}_{Tj} : \widehat{\Sigma}_{Tj}^{-1/2}(\hat{\theta}_{Tj} - \theta_j^0) \to_d \mathcal{N}(0, I)$ as $T \to \infty$ for all $j = [\tau T]$ with some fixed $\tau \in (0, 1)$, with $\mathcal{N}(0, I)$ denoting the multivariate standard normal distribution.

The aim of this paper is to provide a Bayesian treatment to this estimation problem. Note that,

---

[1]The reason why the term $\sum_{t=1}^{T} w_{jt} l_t \left(y_t | y^{t-1}, \theta_t\right)$ can be replaced by $\sum_{t=1}^{T} w_{jt} l_t \left(y_t | y^{t-1}, \theta_j\right)$ in (1) is that $\sum_{t=1}^{T} w_{jt} \left[l_t \left(y_t | y^{t-1}, \theta_t\right) - l_t \left(y_t | y^{t-1}, \theta_j\right)\right]$ is asymptotically negligible; see Giraitis et al. (2014) for details. Intuitively, if j is close to t, $l_t \left(y_t | y^{t-1}, \theta_j\right)$ is close to $l_t \left(y_t | y^{t-1}, \theta_t\right)$ because of continuity of $l(\cdot)$ and conditions (i) and (ii) ensuring that θ_j is close to θ_t. On the other hand, if j is far from t, the weights w_{jt} are close to zero.

[2]While the results of Giraitis et al. (2016) apply for the likelihood function of a Tobit model, the result of Proposition 1 of the current paper applies to any likelihood function, as long as Assumptions 1-5 below are satisfied.

while $\hat{\theta}_{Tj}$ is a consistent extremum estimator, it is not a maximum likelihood estimator because $\ell_{Tj}(\theta_j)$ is not a log-likelihood function due to the kernel weights w_{jt}. Let $\pi_j(\theta_j)$ denote a prior density for θ_j at time j, assumed to be strictly positive and continuous over a compact parameter space Θ. We can combine this prior density with the objective function in (1) to obtain

$$p_{Tj}(\theta_j) = \frac{\pi_j(\theta_j) \exp(\ell_{Tj}(\theta_j))}{\int_\Theta \pi_j(\theta) \exp(\ell_{Tj}(\theta))d\theta}. \tag{4}$$

$p_{Tj}(\theta_j)$ is a proper density for θ_j and can be interpreted as an update to the prior beliefs about θ_j after observing the data. In this sense, $p_{Tj}(\theta_j)$ can be referred to as a quasi-Bayesian (or Laplace type) posterior density based on statistical learning. Properties of such quasi-posterior densities and the resulting estimators[3] have been studied by Chernozhukov and Hong (2003) and Tian, Liu and Wei (2007). In particular, Chernozhukov and Hong (2003) give conditions under which $p_{Tj}(\theta_j)$ can be asymptotically approximated (in total variation of moments norm) by a normal density. As explained in Chernozhukov and Hong (2003), in order for $p_{Tj}(\theta_j)$ to be a valid approximation of a limit distribution for θ_j suitable for confidence interval construction, a generalised information matrix equality must apply for an objective function f that produces an extremum estimator $\hat{\theta}_{Tj}$:

$$\lim_{T \to \infty} \mathbb{E}\left[\nabla f\left(\theta_j^0\right) \nabla f\left(\theta_j^0\right)'\right] = \lim_{T \to \infty}\left[-\mathbb{E}\mathcal{H}\left(f\left(\theta_j^0\right)\right)\right]^{-1} \tag{5}$$

where $\nabla f = \partial f/\partial\theta$ and $\mathcal{H}(f) = \partial^2 f/\partial\theta\partial\theta'$ denote the gradient vector and Hessian matrix of a function $f(\theta)$ respectively. Since this asymptotic equivalence does not hold for the local likelihood function ℓ_{Tj} of Giraitis et al. (2016), we propose a modification that re-scales ℓ_{Tj} by employing a different weighting scheme:

$$\varphi_{Tj}(\theta_j) := \sum_{t=1}^{T} \vartheta_{jt} l_t\left(y_t|y^{t-1}, \theta_j\right) \quad \vartheta_{jt} := \varkappa_{Tj} w_{jt} \quad \text{for } j, t \in \{1, ..., T\} \tag{6}$$

where l_t is the log-likelihood function that appears in (1), and w_{jt} and $\varkappa_{Tj}$ are defined in (2) and (3) respectively. It is clear that the objective functions $\varphi_{Tj}(\theta_j)$ and $\ell_{Tj}(\theta_j)$ give rise to the same maximiser $\hat{\theta}_{Tj}$. However, unlike the case $f = \ell_{Tj}$, the generalised information matrix equality (5) holds for the objective function $f = \varphi_{Tj}$, thereby producing a valid posterior variance. The derivation of the asymptotic results based on objective function (6) is summarised by Proposition

---

[3]A quasi-Bayesian estimator can be defined as the minimiser of some expected loss function; for example, the minimiser of the quadratic loss function is the posterior mean $\int_\Theta \theta_j p_j(\theta_j|Y)d\theta_j$.

1 below, the proof of which can be found in Section 1.1 of the Supplementary Appendix.

Assumptions. For all $j = [\tau T]$ with some fixed $\tau \in (0, 1)$:

1. The parameter space Θ is a compact subset of $\mathbb{R}^{\dim(\theta_j)}$ and $\theta_j^0 \in int(\Theta)$.

2. For any $\delta > 0$, there exists $\lambda > 0$ such that
$$\liminf_{T \to \infty} P \left\{ \sup_{\|\theta_j - \theta_j^0\| > \delta} \frac{1}{\varkappa_{Tj}} \left(\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) \right) \leq -\lambda \right\} = 1.$$

3. For θ_j in an open neighbourhood of θ_j^0, $\varphi_{Tj}(.)$ admits the following expansion
$$\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) = (\theta_j - \theta_j^0)' \nabla \varphi_{Tj}(\theta_j^0) - \frac{1}{2}(\theta_j - \theta_j^0)' \varkappa_{Tj} J_{Tj}(\theta_j^0)(\theta_j - \theta_j^0) + R_{Tj}(\theta_j)$$

where the main components

$$\nabla \varphi_{Tj}(\theta_j) := \frac{\partial \varphi_{Tj}(\theta_j)}{\partial \theta_j} \quad \text{and} \quad J_{Tj}(\theta_j) := -\frac{1}{\varkappa_{Tj}} \mathbb{E} \mathcal{H} \left(\varphi_{Tj}(\theta_j) \right), \tag{7}$$

and the remainder $R_{Tj}(.)$ satisfy:

3a. $-J_{Tj}(\theta_j^0)^{1/2} \nabla \varphi_{Tj}(\theta_j^0) / \sqrt{\varkappa_{Tj}} \to_d \mathcal{N}(0, I)$ as $T \to \infty$.

3b. The maximal and minimal eigenvalue of $J_{Tj}(\theta_j^0)$ satisfy: $\lambda^* = \limsup_{T \to \infty} \lambda_{\max}(J_{Tj}(\theta_j^0)) < \infty$, $\lambda_* = \liminf_{T \to \infty} \lambda_{\min}(J_{Tj}(\theta_j^0)) > 0$.

3c. For every $\varepsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ and $M > 0$ such that
$$\limsup_{T \to \infty} P \left\{ \sup_{M/\sqrt{\varkappa_{Tj}} \leq \|\theta - \theta_j^0\| \leq \delta} \frac{\|R_{Tj}(\theta)\|}{\varkappa_{Tj} \|\theta - \theta_j^0\|^2} > \eta \right\} \leq \varepsilon,$$
$$\limsup_{T \to \infty} P \left\{ \sup_{\|\theta - \theta_j^0\| \leq M/\sqrt{\varkappa_{Tj}}} \|R_{Tj}(\theta)\| > \varepsilon \right\} = 0.$$

4. The prior density $\pi_j(.)$ is strictly positive and Lipschitz continuous function over Θ.

5. The time variation in the true parameters θ_j^0 satisfies one of conditions (i) or (ii).

Proposition 1. *Denote by $p_\infty(\cdot)$ the probability density of the $\mathcal{N}\left(0, J_{Tj}(\theta_j^0)^{-1}\right)$ distribution, where $J_{Tj}(.)$ is defined in (7). Consider the following family of random vectors:*
$$\mathbb{H}_{Tj} := \left\{ h_{Tj} \equiv \sqrt{\varkappa_{Tj}} \left(\theta_j - \theta_j^0 \right) - J_{Tj}(\theta_j^0)^{-1} \nabla \varphi_{Tj}(\theta_j^0) / \sqrt{\varkappa_{Tj}} : \theta_j \in \Theta \right\}.$$
Then, under Assumptions 1-5, for all $j = [\tau T]$ with some fixed $\tau \in (0, 1)$, the quasi-posterior density of h_{Tj}, $p_{Tj}(h_{Tj})$, based on the objective function (6), is asymptotically approximated by $p_\infty(h_{Tj})$ in total variation of moments norm: $\int_{\mathbb{H}_{Tj}} (1 + \|h\|^\alpha) |p_{Tj}(h) - p_\infty(h)| \, dh \to_p 0$ as $T \to \infty$ for all $\alpha > 0$.

Remarks

1. The result of Proposition 1 is general and applies to any local likelihood objective function. Assumptions 2 and 3 are high level assumptions concerning, respectively, identification and quadratic

approximation of the objective function; they provide a time varying analogue of Assumptions 3 and 4 in Chernozhukov and Hong (2003). We replace Chernozhukov and Hong (2003)'s asymptotic analysis based on $\sqrt{T}$-neighbourhoods of the true (in their analysis fixed) parameter θ^0 to general $\varkappa_{Tj}^{1/2}$-neighbourhoods of θ_j^0 (now indexed by time) with the sequence $\varkappa_{Tj}$ defined in (3). Assumption 4 explicitly allows the prior density π_j to vary over time by imposing a Lipschitz continuity requirement: this is a stronger assumption than the (time invariant) prior density assumed in Assumption 2iv) in Chernozhukov and Hong (2003) and is required to accommodate time variation. Any continuously differentiable density function for $\theta_j \in \Theta$ satisfies Assumption 4. Finally, Assumption 5 imposes structure on the types of time variation the method can handle, the key requirement being consistent estimation in the frequentist setup of Giraitis et al. (2014, 2016, 2018), see the discussion in the beginning of Section 2. The high level Assumptions 2 and 3 require verification on a case-by-case basis according to the model at hand. In Remark 2 below, we show how these assumptions can be verified in the case of a linear Gaussian local likelihood function studied in Sections 3-5 of this paper.

2. For the linear Gaussian VAR models studied in Sections 3-4 and applied to various economic problems and forecasting in Section 5, the local likelihood function in (6) admits the closed form (9) derived in Section 3 below. For this objective function, Assumption 2 follows immediately by Lemma 4a of the Supplementary Appendix of Giraitis et al. (2016). Assumption 3c requires that the remainder term $R_{Tj}(\theta_j)$ of the quadratic approximation of the objective function is asymptotically negligible. By Lemma 2 in Chernozhukov and Hong (2003), Assumption 3c is satisfied for any twice continuously differentiable objective function as long as a uniform law of large numbers applies to the Hessian matrix. This is indeed the case for the Gaussian local likelihood objective functions studied in Sections 3-5 of this paper: twice continuous differentiability is immediate and a uniform law of large numbers for the Hessian applies by virtue of Lemma 4b in Giraitis et al. (2016), thereby ensuring the validity of Assumption 3c. Additionally, the asymptotic normality result in Theorem 1 of Giraitis et al. (2016) guarantees the fulfillment of Assumption 3a. The above considerations ensure the validity of Assumptions 2 and 3 in the linear Gaussian case; hence, the asymptotic result of Proposition 1 applies to Propositions 2-3 of Section 3 below.

3. By consistency of $\hat{\theta}_{Tj}$, $\frac{1}{\varkappa_{Tj}} \nabla \varphi_{Tj}(\theta_j^0) \to_p 0$ as $T \to \infty$. Proposition 1 therefore implies that when the sample size T is large, draws from $p_{Tj}(\theta_j)$ have the following approximate distribution $\theta_j^{sim} \sim \mathcal{N}\left(\theta_j^0, \left[\varkappa_{Tj} J_{Tj}(\theta_j^0)\right]^{-1}\right)$. Since $\varkappa_{Tj} \to \infty$ for all j, we conclude that the mean, median and mode of the quasi-posterior distribution $p_{Tj}(\theta_j)$ are also consistent. Moreover, confidence intervals

constructed using the variance of $p_{Tj}(\theta_j)$ provide asymptotically valid inference and the quasi-posterior has the same asymptotic variance as the extremum estimator of Giraitis et al. (2016).

4. The violation of the information equality (5) by the objective function ℓ_{Tj} of Giraitis et al. (2016) is irrelevant for the frequentist estimators studied in that paper; it only becomes an issue after augmentation of the local likelihood ℓ_{Tj} by a prior density. To provide an intuition of why normalisation of the weights by $\varkappa_{Tj}$ is required in (6), note that, if the weights w_{jt} satisfy $\sum_{t=1}^{T} w_{jt} = 1$ as in Giraitis et al. (2016), the local likelihood would not dominate the prior as $T \to \infty$. For the prior to vanish, we require that the sum of the weights w_{jt} increases with the sample size; and in order to obtain the same rate of convergence as in Giraitis et al. (2016), the weights ϑ_{jt} need to sum to $\varkappa_{Tj}$.

3 QBLL Inference in a Linear Gaussian Multivariate Setting

3.1 Quasi-posterior distribution

In this section we derive closed form Normal-Wishart expressions for the quasi-posterior distribution of a VAR model with time varying parameters.

Suppose that y_t is generated by a time varying parameter (TVP) VAR process of lag order k given by:

$$y_t = B_{0t} + \sum_{p=1}^{k} B_{pt} y_{t-p} + \varepsilon_t, \quad \varepsilon_t = R_t^{-1/2} \eta_t, \quad \eta_t \sim NID(0, I_M) \tag{8}$$

where B_{0t} is a vector of time varying intercepts, B_{pt} are time varying autoregressive matrices and R_t^{-1} is a positive definite time varying covariance matrix. We assume that all roots of the polynomial $\psi(z) = \det\left(I_M - \sum_{p=1}^{k} z^p B_{pt}\right)$ lie outside the unit circle. Letting $x_t = (1, y'_{t-1}, ..., y'_{t-k})$ and $B_t = (B_{0t}, B_{1t}, ..., B_{kt})$, the model (8) can be written as $y_t = (I_M \otimes x_t)\beta_t + R_t^{-1/2}\eta_t$, where $\beta_t = vec(B'_t)$. Further, define the matrices $Y = (y_1, ..., y_T)'$, $E = (\varepsilon_1, ..., \varepsilon_T)'$, $X = (x'_1, ..., x'_T)'$ and denote their vectorised forms by $y = vec(Y)$ and $\varepsilon = vec(E)$. Employing the normalisation of the weights proposed in the previous section in (6), the weighted likelihood for the VAR(k) model (8) at each point in time $j \in \{1..., T\}$ is given by

$$
\begin{aligned}
L_j(y|\beta_j, R_j, X) &= (2\pi)^{-M\varkappa_{Tj}/2} |R_j|^{\varkappa_{Tj}/2} e^{-\frac{1}{2}\sum_{t=1}^{T} \vartheta_{jt}(y_t - (I_M \otimes x_t)\beta_j)' R_j(y_t - (I_M \otimes x_t)\beta_j)} \\
&\propto |R_j|^{tr(D_j)/2} \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta_j)'(R_j \otimes D_j)(y - (I_M \otimes X)\beta_j)\right\} \quad (9)
\end{aligned}
$$

where $\varkappa_{Tj}$ is defined in (3), the kernel weights ϑ_{jt} are defined in (6) and $D_j = diag(\vartheta_{j1}, ..., \vartheta_{jT})$.

Next, assume that β_j and R_j have Normal-Wishart prior distribution for $j \in \{1, ..., T\}$:

$$\beta_j | R_j \sim \mathcal{N}\left(\beta_{0j}, (R_j \otimes \kappa_{0j})^{-1}\right), \quad R_j \sim \mathcal{W}(\alpha_{0j}, \gamma_{0j}^{-1}) \tag{10}$$

where β_{0j} is a vector of prior means, κ_{0j} is a positive definite matrix, α_{0j} is a scalar scale parameter of the Wishart distribution and γ_{0j} is a positive definite matrix.

Proposition 2. *The weighted likelihood L_j in (9) combined with the prior in (10) results to a Normal-Wishart quasi-posterior distribution for β_j and R_j:*

$$\beta_j | R_j, X, Y \sim \mathcal{N}\left(\widetilde{\beta}_j, (R_j \otimes \widetilde{\kappa}_j)^{-1}\right), \quad R_j \sim \mathcal{W}(\widetilde{\alpha}_j, \widetilde{\gamma}_j^{-1}) \tag{11}$$

for $j \in \{1, ..., T\}$ with posterior parameters:

$$\widetilde{\beta}_j = \left(I_M \otimes \widetilde{\kappa}_j^{-1}\right)\left[(I_M \otimes X'D_jX)\hat{\beta}_j + (I_M \otimes \kappa_{0j})\beta_{0j}\right], \tag{12}$$

$$\widetilde{\kappa}_j = \kappa_{0j} + X'D_jX, \quad \widetilde{\alpha}_j = \alpha_{0j} + \sum_{t=1}^{T} \vartheta_{jt}, \quad \widetilde{\gamma}_j = \gamma_{0j} + Y'D_jY + B_{0j}\kappa_{0j}B_{0j}' - \widetilde{B}_j\widetilde{\kappa}_j\widetilde{B}_j',$$

where $\hat{\beta}_j = (I_M \otimes X'D_jX)^{-1}(I_M \otimes X'D_j)y$ is the local likelihood estimator for β_j.

The marginal distribution of the parameter vector β_j can be obtained by integrating the quasi-posterior distribution $p(\beta_j, R_j | Y, X)$ obtained in Proposition 2 over the $M(M+1)/2$ distinct elements of the symmetric matrix R_j.

Proposition 3. *The marginal distribution of β_j is given by a multivariate non-standardised t-distribution with $\widetilde{\alpha}_j - Mk$ degrees of freedom: $\beta_j | Y, X \sim T_{\widetilde{\alpha}_j - Mk}\left(\widetilde{\beta}_j, \frac{\widetilde{\gamma}_j \otimes \widetilde{\kappa}_j^{-1}}{\widetilde{\alpha}_j - Mk - 2}\right)$, where $\widetilde{\alpha}_j$, $\widetilde{\beta}_j$, $\widetilde{\gamma}_j$ and $\widetilde{\kappa}_j$ are defined in Proposition 2.*

The proofs of Propositions 2-3 as well as some additional closed form results can be found in Sections 1.2-1.6 of the Supplementary Appendix.

Remarks

1. The quasi-posterior derived in Proposition 2 shares similarity with the Normal-Wishart conjugate posterior result in the Bayesian VAR literature. However, the QBLL approach is adapted to include the kernel weights ϑ_{jt} and provides posterior distributions for each point in time j, thereby accommodating time variation in the parameters. Note that the result in Proposition 1 applies to a

multivariate Gaussian system of a general form; however, since the focus of this paper is on BVAR models, the paper restricts attention to VAR systems.

2. While the model in (8) is of parametric form, the time variation in the parameters is dealt with nonparametrically: the sequence of parameters needs only satisfy one of the "slow drift" conditions (i) or (ii), which encompass a wide class of processes without the need to impose a specific modelling restrictions on the parameters.

3. The QBLL approach provides a framework for accommodating prior beliefs on the presence of time variation in the parameters, since the prior parameters are allowed to vary over time.

4. The joint quasi-posterior of $\beta_1, \beta_2, ..., \beta_T$ and $R_1, R_2, ..., R_T$ is given by

$$p(\beta_{1:T}, R_{1:T}|Y, X) = p(\beta_1, R_1|Y, X) p(\beta_2, R_2|Y, X) ... p(\beta_T, R_T|Y, X)$$

because for any periods i and j the pairs (β_i, R_i) and (β_j, R_j) are conditionally independent given the entire dataset Y, X, since $\sigma(Y, X) = \sigma(X, Y, \theta_k)$ for any period k, with $\sigma(\cdot)$ denoting a generated sigma algebra.

5. Quasi-Bayesian point estimates for β_j and R_j^{-1} can be obtained by using the quasi-posterior means in Proposition 2. For β_j, the posterior mean is given by $\widetilde{\beta}_j$ in (12). For the covariance matrix R_j^{-1}, the Wishart property of R_j implies an inverse Wishart $\mathcal{IW}(\tilde{\alpha}_j, \tilde{\gamma}_j)$ distribution for R_j^{-1}, with posterior mean given by $\tilde{\gamma}_j/(\tilde{\alpha}_j - M - 1)$ in the notation of Proposition 2.

6. In view of the modification of the kernel weights in (6), the quasi-Bayesian estimators in Remark 5 are asymptotically equivalent to the local likelihood estimator of Giraitis et al. (2016) (which in the linear Gaussian model also coincides with the estimator in Giraitis et al. (2014)) and thereby inherit the latter's consistency and asymptotic normality properties. In particular, $\widetilde{\beta}_j$ is a weighted average of the prior mean, β_{0j}, and the local likelihood estimator $\hat{\beta}_j$ with relative weights given by $I_M \otimes (\kappa_{0j} + X'D_jX)^{-1}\kappa_{0j}$ and $I_M \otimes (\kappa_{0j} + X'D_jX)^{-1}X'D_jX$ respectively. The former expression depends positively on the prior tightness κ_{0j}^{-1} and converges to zero as $T \to \infty$; the latter expression converges to unity. Convergence of both expressions (and hence asymptotic negligibility of the priors) applies because the modified[4] kernel weights ϑ_{jt} ensure that $\text{plim}_{T \to \infty} \lambda_{\min}(X'D_jX) = \infty$. Similarly, the posterior mean for R_j^{-1} is asymptotically equivalent to the local likelihood estimator of Giraitis et al. (2014, 2016, 2018).

7. Despite the asymptotic equivalence of the quasi-Bayesian and frequentist kernel-type estimators discussed in Remark 6, the prior distribution may have a considerable small sample effect: a

---

[4]A naïve approach whereby the objective functions of Giraitis et al. (2014, 2016, 2018) are augmented with a prior, without the normalisation of the kernel weight proposed in the previous section in (6), would give rise to inconsistent quasi-Bayesian estimators since the prior will not vanish as $T \to \infty$.

suitable choice of prior can improve the finite sample performance of the frequentist estimator and can help avoid over-fitting in over-parameterised VAR systems. Moreover, the prior plays a crucial role in large-dimensional VAR systems, where the local likelihood estimator may be undefined due to lack of invertibility of $X'D_jX$. In such cases, the presence of the prior ensures feasibility of the QBLL estimator.

8. Frequentist penalised regression estimators such as Ridge, Lasso or Elastic net arise as special cases of the posterior mode of our quasi-Bayesian posterior distribution with Normal, Laplace or mixture prior for β_j respectively.

3.2 Discussion

We provide a comparison of the relative merits of the proposed QBLL approach to the widely used state space approaches to the estimation of the time varying parameters β_j and R_j^{-1}. The most serious limitation of state space methodology is the inability to accommodate large dimensional VAR systems, and we provide an additional comparison in Section 3.2.2 of our QBLL approach to alternative methods that are able to deal with large dimensions.

3.2.1 Comparison to state space models

Writing a TVP VAR model in state space involves adding the drifting parameters to a state vector of latent variables. This requires the specification of a stochastic process for modelling these unobserved parameters, with a random walk process being the most common assumption in the literature (see for example, Cogley and Sargent (2002, 2005), Primiceri (2005), Mumtaz and Surico (2009), Cogley et al. (2010) and Clark (2012)). This assumption is convenient as it is the simplest way to induce persistence in the parameters and, at the same time, considerably reduces the number of additional coefficients added to each state equation. However, the Kalman filter is not optimal under misspecification of the state space and inaccurately specifying a random walk state equation can invalidate inference, even asymptotically. The finite sample distortions resulting from such misspecification can be as severe as coverage rates of 30% for a nominal of 95%, as demonstrated in the Monte Carlo exercise in Section 4.

Another issue that arises when modelling the drifting covariance matrix R_j^{-1} in a state space setup is the potential loss of the positive definite property of this matrix, if the elements of R_j^{-1} were modelled as drifting stochastic processes. To address this issue, Cogley and Sargent (2005), Primiceri (2005) and Cogley et al. (2010) diagonalise the covariance matrix using a Cholesky decomposition and then assume that the diagonal elements follow a random walk in logarithms, thus ensuring they remain positive. This requires additional modelling assumptions relating to the sto-

13

chastic properties of the elements of lower triangular matrix in the Cholesky decomposition. In addition, it implies that the ordering of the variables in the VAR matters for inference. Such triangularisation of the covariance matrix is useful for conducting structural analysis and Canova and Perez Forero (2015) show how to estimate time varying identifying restrictions that are not necessarily just-identified and recursive. However, in models where structural shock analysis is not required and the VAR model is used for forecasting, diagonalisation has only been employed as means to facilitate estimation of the drifting volatilities (Cogley and Sargent (2005), Clark (2012), Carriero, Clark and Marcellino (2016)) and has the undesirable effect of making the model-implied forecasts depend on the ordering of the variables in the system. The QBLL approach permits direct estimation of R_j^{-1}, which has an inverted-Wishart posterior density and hence remains symmetric and positive definite at each point in time, thus making diagonalisation redundant. When applied to structural shock analysis, the QBLL estimator is informative on the presence of time variation in the identifying restrictions, this information being an outcome of the estimation procedure rather than a maintained assumption. In addition, when used for forecasting, the QBLL approach delivers time variation in the covariance matrix of the reduced form VAR model regardless of the ordering of the variables in the system.

The state equation has sometimes been given a prior density interpretation. For example, Cogley and Sargent (2005) assume $p(\beta_t) = \mathcal{N}(\beta_{t-1}, Q)$ which is equivalent to a random walk equation for β_t. This might appear to coincide with the QBLL estimator whenever the prior in (10) is of the form $\beta_j | R_j \sim \mathcal{N}\left(\beta_{j-1}, (R_j \otimes \kappa_{0j})^{-1}\right)$. However, as illustrated in the Monte Carlo, when the state equation is misspecified, invalid inference can occur even asymptotically. This suggests that the state equation is not a proper prior in the sense that it does not vanish asymptotically. On the other hand, as shown in Proposition 1, the prior of the QBLL estimator has no asymptotic effect on the posterior. In addition, state space models can run into issues such as non-stationary draws from the time varying autoregressive matrices. Such difficulties can be overcome if the drifting parameters could be directly 'shrunk' at each point in time, as is the case with the proposed QBLL estimator.

Perhaps the most important limitation to the practical application of state space methodology is their inability to accommodate large dimensional models. The size and complexity of the state space increases substantially with the VAR dimension, since an extra state equation is required for each parameter as well as an additional shock and additional coefficients guiding the process[5]. As a result,

---

[5]In a state space setting, an M-dimensional TVP VAR (k) with stochastic volatility requires the addition of

state space models suffer from dimensionality problems and their application to the estimation of TVP VAR models is limited to a model of three to four variables. Additional complexity in the estimation of state space models arises from the numerical approximation of the joint posterior density conducted by MCMC algorithms. On the other hand, the proposed QBLL methodology admits a closed form quasi-posterior density, facilitating estimation of VAR systems of size of over 80 variables[6] in a little over a minute of computation time, as demonstrated in Section 5.2.

3.2.2 Comparison to alternative methods for large TVP VARs

To address the dimensionality problems of state space models, Koop and Korobilis (2013) employ a 'forgetting factor' approach, which involves discounting past observations in the Kalman filter prediction step. Their approach provides a remedy to the dimensionality problem but remains susceptible to the modelling limitations of state space methods discussed in the previous section; e.g. lack of robustness to misspecification of the state equations. Carriero et al. (2016) propose an algorithm, which is designed to draw from the conditional posterior of the autoregressive parameters equation-by-equation, hence reducing the computational complexity of state space models in larger systems. The methodology requires a Cholesky triangularisation scheme and assumes a VAR model with time invariant parameters and stochastic volatility, hence it remains prone to lack of robustness due to i) state equation parametric assumptions, and ii) the triangular matrix imposing that the order of the variables in the system matters. The papers of Kapetanios and Zikes (2017) and Kapetanios, Marcellino and Venditti (2015) also propose a solution to the dimensionality problem, employing penalised Lasso and Ridge procedures respectively, based on the frequentist kernel estimator of Giraitis et al. (2014). As explained in Remark 8 in Section 3, the posterior mode of the QBLL estimator proposed in this paper can encompass both penalised estimators with the appropriate choice of prior density. Moreover, the treatment of the time varying parameters in Kapetanios and Zikes (2017) and Kapetanios et al. (2015) is frequentist and does not allow for: (i) posterior distributions varying over time; (ii) a mechanism for taking into account prior beliefs about the volatilities. The QBLL approach produces a closed form expression for the posterior density in (11) and can incorporate prior beliefs about R_j^{-1} in the form of the parameters α_{0j} and γ_{0j}. An additional advantage of the methodology of this paper is that the Bayesian analysis provides a framework for the construction of MCMC algorithms which enable the estimation of

---

$M(3/2 + M(k + 1/2))$ state equations, so for instance, a four variable TVP VAR(2) requires 46 state equations.

[6]A VAR(1) model with 80 variables with time variation in the parameters and the covariance matrix would require the use of 9720 state equations. Even in a setup with a fixed parameter VAR with stochastic volatility, Carriero et al. (2016) show that the space space model requires M^3 elementary operations per a single draw of the MCMC.

mixtures of time varying and time invariant parameters.

3.3   A Gibbs sampling approach

There are cases when the researcher would like to consider model (8) with time invariant autoregressive coefficients β and time varying variances R_t^{-1}. Such models have been proposed by Primiceri (2005) and Sims and Zha (2006), and have become popular for being able to fit well macroeconomic data when the sample considered contains periods characterised by changing volatility; for example, data containing the high volatility periods of the oil crises followed by the low volatility periods of the Great Moderation. Alternatively, the researcher might want to keep the variances of the errors fixed over time and allow for time varying autoregressive coefficients, as in Cogley and Sargent (2002). This approach also has merits: an AR model with drifting autoregressive coefficients is a flexible framework that can fit a range of nonlinearities. Such models are popular and widely used in other disciplines such as image and signal processing (see for example, Kitagawa (1996), Abramovich, Spencer and Turley (2007)). Another approach, useful in some applications, assumes most parameters to be fixed but one or two, interesting from economic perspective, are allowed to vary over time.

It is therefore desirable to develop a framework for inference that encompasses all the above scenaria by allowing for mixtures of time varying and time invariant parameters. To this end, we utilise the Bayesian theory developed in this paper in order to design several Gibbs algorithms that can produce approximate joint posterior distributions for such mixtures. These cases are not covered by the estimators of Giraitis et al. (2014, 2016, 2018) and are only covered by state space models at the cost of possible state equation misspecification and VAR dimensionality restrictions, as discussed in Section 3.2.1. The Gibbs algorithms are based on the idea that: (i) when conditioning on a draw from the time invariant covariance or autoregressive matrix, the conditional quasi-posterior distributions of the remaining time varying parameters are available in closed form and provided in Proposition 4 and 5 in Section 1.4 of the Supplementary Appendix; (ii) given a draw from these closed form expressions, the posterior of the time invariant covariance or autoregressive matrix is of standard form. Such models permits the use of Gibbs sampling algorithms to recursively draw from the known conditional quasi-posteriors in order to provide an approximation of the unknown joint quasi-posterior of the mixtures of time varying and time invariant parameters. A detailed description of the resulting algorithms in the case of homoscedastic BVAR model with time varying parameters and heteroscedastic BVAR model with time invariant parameters can be found in Sections 2.1 and 2.2 of the Supplementary Appendix respectively. The Appendix also contains

an additional algorithm for a time varying structural BVAR model.

The proposed algorithms perform well with different parameter processes and allow the researcher to remain agnostic about the drifting parameters' data generating process.

4   Monte Carlo

In this section, we design a Monte Carlo exercise to study the finite sample properties of the QBLL estimators introduced in this paper and how they compare to the alternative state space approach. We limit our attention to a two dimensional VAR(1) model with no intercept term. We simulate data using four different data generating processes (DGPs) and estimate the models we want to compare based on these simulated samples. In DGPs II-III, the state space model is expected to perform better than the semi-parametric QBLL method, since it employs the correct parametric specification for the state equation and hence uses more information in the estimation procedure. However, we demonstrate that, even in these unfavourable cases, the QBLL estimator has comparable finite sample properties and, crucially, is always consistent, implied by the monotonically decreasing errors with the increasing sample sizes. In DGP IV, we illustrate how misspecifying the state equation can have severe consequences and lead to invalid inference with root mean square error (RMSE) which does not vanish asymptotically for the parametric state space models; the QBLL approach, on the other hand, remains consistent and its finite sample performance is unaffected. Details on the Monte Carlo design can be found in Section 3.1 of the Supplementary Appendix.

DGP I. The first DGP we consider is a model with fixed parameters and fixed volatility:

$$
\begin{aligned}
y_t &= B y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Psi), \quad y_0 = 0, \\
B &= \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.6 \end{bmatrix}, \ \Psi = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}.
\end{aligned}
\tag{13}
$$

Our motivation is to assess how well models featuring time varying parameters can fit a simple time invariant model. Table 1 presents the bias, RMSE and coverage rates for the QBLL estimator featuring time varying parameters and volatility (QBLL-TV) and a state space model, with drifting parameters and volatilities, referred to as SS-TV. It is clear from Table 1 that both models deliver valid inference. For small samples, the state space approach performs better in terms of RMSEs for the parameters and worse for the volatility, but when the sample size increases, both models

perform similarly. The confidence intervals implied by the QBLL approach are close to the nominal 95% for all sample sizes. On the other hand, the confidence intervals of the state space model were found to be very sensitive to the prior tightness on the standard deviation in the random walk state equations. In particular, we found that increasing the prior tightness delivers worse coverage rates, while loosening the prior, which is a way to 'let the data speak', considerably increases the difficulty of obtaining a stable draw from the autoregressive matrix of the VAR and therefore the time required to estimate the model.

DGP I							
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
T=100	QBLL-TV	-0.0307	0.1698	0.9326	-0.0105	0.0605	0.9292
	SS-TV	-0.0167	0.1191	0.9757	0.0131	0.0701	0.9679
T=500	QBLL-TV	-0.0134	0.1033	0.9457	-0.0030	0.0410	0.9383
	SS-TV	-0.0073	0.0696	0.9890	0.0090	0.0459	0.9779
T=1000	QBLL-TV	-0.0126	0.0874	0.9399	-0.0051	0.0341	0.9424
	SS-TV	-0.0080	0.0602	0.9858	0.0039	0.0391	0.9677

Table 1. Bias, RMSEs and coverage rates of models based on DGP I for 100, 500 and 1000 observations respectively.

DGP II. The second DGP is given by a model with fixed autoregressive matrix as in (13) and time varying volatility: $\varepsilon_t \sim NID(0, \Sigma_t)$, $\Sigma_t = H_t^{-1} \Psi_t \left(H_t^{-1} \right)'$, with H_t a triangular matrix with ones on the main diagonal and bottom left element h_{21}, Ψ_t is a diagonal matrix, with diagonal elements, φ_{11} and φ_{22}, and $h_{21,t} = \frac{\sum_{i=1}^{t} \xi_i}{\sqrt{t}}$, $\xi_i \sim NID(0, 0.1^2)$, $\log \varphi_{kk,t} = \frac{\sum_{i=1}^{t} v_i^k}{\sqrt{t}}$, $v_i^k \sim NID(0, 0.1^2)$, $k = 1, 2$.

DGP II							
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
	QBLL-TV	-0.0301	0.1532	0.9334	-0.0138	0.2186	0.9328
T=100	SS-TV	-0.0169	0.1065	0.9776	0.0442	0.2458	0.9633
	QBLL-TVV	-0.0073	0.0801	0.9492	0.0232	0.2405	0.9564
	SS-TVV	-0.0076	0.0872	0.9408	0.0884	0.3030	0.9735
	QBLL-TV	-0.0137	0.0940	0.9416	-0.0057	0.1490	0.9401
T=500	SS-TV	-0.0071	0.0535	0.9903	0.0238	0.1684	0.9499
	QBLL-TVV	-0.0023	0.0345	0.9492	0.0114	0.1546	0.9491
	SS-TVV	-0.0026	0.0380	0.9417	0.0399	0.2003	0.9605
	QBLL-TV	-0.0110	0.0777	0.9414	-0.0093	0.1254	0.9326
T=1000	SS-TV	-0.0063	0.0444	0.9841	0.0092	0.1463	0.9384
	QBLL-TVV	-0.0028	0.0260	0.9400	0.0048	0.1290	0.9424
	SS-TVV	-0.0029	0.0279	0.9250	0.0203	0.1667	0.9570

Table 2. Bias, RMSEs and coverage rates of models based on DGP II for 100, 500 and 1000 observations respectively.

The processes used for $h_{21,t}$, $\log \varphi_{11,t}$ and $\log \varphi_{22,t}$ resemble random walk processes, but are normalised by $\sqrt{t}$ in order to have finite variance, see Giraitis et al. (2014). In addition to the two models we fit to DGP I: QBLL-TV and SS-TV, we also include a QBLL model with fixed autoregressive parameters and time varying volatility, estimated using Algorithm 2 of Section 2.2

of the Supplementary Appendix (QBLL-TVV) and, similarly, a state space model with constant parameters and stochastic volatility (SS-TVV). Table 2 displays the bias, RMSEs and coverage rate of the different models estimated with data generated with DGP II. It is clear from Table 2 that both QBLL-TV and SS-TV approaches perform well, with the state space model delivering: (i) smaller RMSEs for the autoregressive parameters and, (ii) larger RMSEs for the volatility processes. Additionally, the two models featuring fixed autoregressive coefficients and drifting volatility: QBLL-TVV and SS-TVV perform better than QBLL-TV and SS-TV respectively, particularly for the autoregressive parameters. This is to be expected, since both exploit the additional information of parameter invariance in the estimation procedure.

		DGP III					
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
T=100	QBLL-TV	-0.0046	0.1703	0.9473	-0.0275	0.2191	0.9350
	SS-TV	-0.0059	0.1295	0.9716	0.0366	0.2512	0.9568
T=500	QBLL-TV	-0.0001	0.1129	0.9449	-0.0106	0.1527	0.9423
	SS-TV	0.0004	0.0727	0.9775	0.0212	0.1741	0.9567
T=1000	QBLL-TV	-0.0022	0.0911	0.9500	-0.0082	0.1247	0.9447
	SS-TV	-0.0018	0.0555	0.9849	0.0155	0.1453	0.9540

Table 3. Bias, RMSEs and coverage rates of models based on DGP III for 100, 500 and 1000 observations respectively.

DGP III. The third DGP is a model featuring both drifting parameters and volatility

$$y_t = B_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Sigma_t), \quad y_0 = 0. \tag{14}$$

We use the same process for the volatility Σ_t as in DGP II. In addition, the autoregressive parameters are generated as bounded random walks, subject to stability at each point in time: $B_{km,t} = \frac{\sum_{i=1}^{t} \zeta_i}{\sqrt{t}}, \zeta_i \sim NID(0, 0.1^2),$ for $k, m = 1, 2$. Table 3 displays the bias, RMSE and coverage rate of the different models when the data are generated using DGP III. From Table 3, it clear that both the QBLL model and the state space model perform well; the state space model delivers smaller RMSEs for the autoregressive parameters and larger RMSEs for the volatility. The differences between the two models' point estimates narrow when the sample size increases.

DGP IV. The forth DGP is given by model (14), but here the autoregressive parameters and volatility follow the processes

$$\begin{aligned}
B_{km,t} &= 0.25 \sin\left(0.004\pi t\right) + 0.25 \sum_{i=1}^{t} \zeta_i / \sqrt{t}, & \zeta_i &\sim NID(0, 0.3^2), \quad \text{for} \quad m, k = 1, 2, \\
h_{12,t} &= 0.5 \sin\left(0.004\pi t\right) + 0.5 \sum_{i=1}^{t} \xi_i / \sqrt{t}, & \xi_i &\sim NID(0, 0.3^2), \\
\log \varphi_{kk,t} &= 0.5 \sin\left(0.004\pi t\right) + 0.5 \sum_{i=1}^{t} v_i^k / \sqrt{t}, & v_i^k &\sim NID(0, 0.3^2), \quad k = 1, 2.
\end{aligned}$$

19

All parameters in DGP IV combine a deterministic process following a sine wave with a stochastic part, generated by the bounded random walk process, used in DGPs II and III. For the estimation of the state space model, if the researcher knows ex-ante that the DGP contains a deterministic time trend, this information can be further added to the state equations. However, we continue using the random walk processes for the parameters, as these are widely used in the literature and the motivation behind our choice is to assess the consequences of using the 'wrong' state equation.

		DGP IV					
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
	QBLL-TV, H=T$^{0.5}$	-0.0088	0.1853	0.9451	-0.0779	0.3657	0.9227
	QBLL-TV, H=T$^{0.45}$	-0.0120	0.2034	0.9456	-0.0904	0.3982	0.9205
T=100	QBLL-TV, H=T$^{0.55}$	-0.0060	0.1694	0.9435	-0.0679	0.3382	0.9238
	SS-TV	-0.0094	0.1332	0.8922	-0.0372	0.3934	0.8040
	F-BVAR	0.0056	0.1303	0.8799	-0.0606	0.3673	0.7306
	QBLL-TV, H=T$^{0.5}$	0.0057	0.1209	0.9470	-0.0204	0.2130	0.9313
	QBLL-TV, H=T$^{0.45}$	0.0041	0.1384	0.9498	-0.0255	0.2402	0.9334
T=500	QBLL-TV, H=T$^{0.55}$	0.0088	0.1074	0.9399	-0.0163	0.1949	0.9202
	SS-TV	0.0302	0.1437	0.5176	-0.0349	0.3282	0.5560
	F-BVAR	0.0304	0.1366	0.3992	0.0384	0.1925	0.3789
	QBLL-TV, H=T$^{0.5}$	0.0092	0.1004	0.9464	-0.0009	0.1833	0.9315
	QBLL-TV, H=T$^{0.45}$	0.0035	0.1174	0.9485	-0.0100	0.2006	0.9418
T=1000	QBLL-TV, H=T$^{0.55}$	0.0202	0.0930	0.9213	0.0104	0.1864	0.8666
	SS-TV	0.0545	0.1812	0.2936	-0.0015	0.3613	0.4918
	F-BVAR	0.1127	0.2244	0.1697	0.1248	0.5559	0.0984

Table 4. Bias, RMSEs and coverage rates of models based on DGP IV for 100, 500 and 1000 observations respectively. The table also contains a comparison of the QBLL estimator with bandwidth parameters T$^{0.45}$, T$^{0.5}$ and T$^{0.55}$.

Table 4 summarises the resulting bias, RMSE and coverage rate of the different models, including a fixed parameter BVAR model (F-BVAR) and two additional QBLL estimators, computed using different values of the bandwidth parameter H, when the sample is simulated using DGP IV. From Table 4 it is clear that both the fixed parameter model and the state space model provide invalid inference. Crucially, the problem does not vanish asymptotically, with RMSEs and coverage rates deteriorating with the sample size: the true autoregressive parameters are outside their 95% confidence intervals more than 70% of time implied by the state space model, and more than 80% of the time implied by the fixed parameter BVAR model for a sample size of 1000. We draw two important conclusions from these results. First, parameter time variation is an issue that cannot be ignored since fixed parameter models in this case fail to deliver valid inference. Second, the state space model also fails to deliver valid inference, as it models the parameter process parametrically through the state equation and hence its statistical properties depend on the validity of the assumptions about the state equation. The nonparametric QBLL estimators on the other hand remain valid in this case. Figure 2 further illustrates this last point by displaying

Figure 2: DGP IV. Typical realisation of the time varying parameters and volatilities
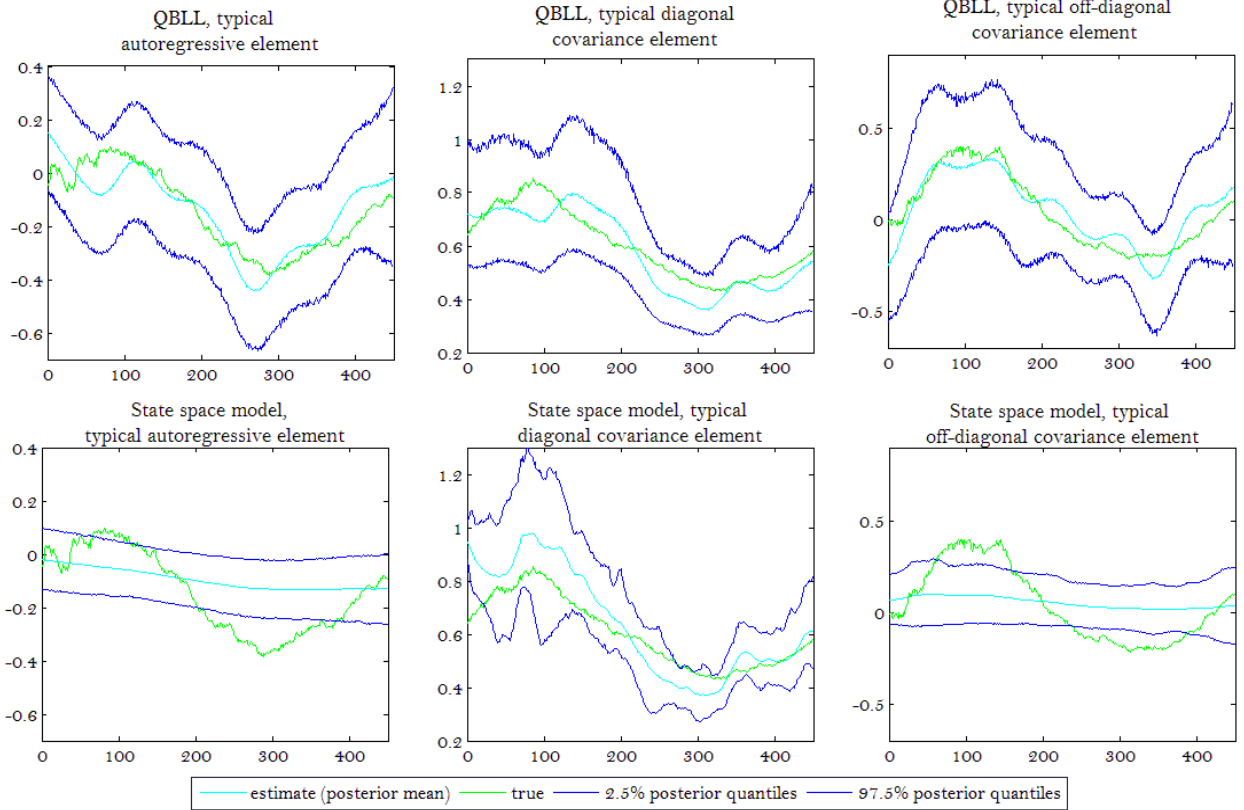
typical realisations of the parameters over time: the state space model in this case does not account well for time variation and provides invalid confidence intervals. Table 4 also compares the QBLL estimator with bandwidth parameter $H = T^{0.5}$ with two alternatives: $H = T^{0.45}$ and $H = T^{0.55}$. From the table, we see that decreasing the bandwidth reduces the bias and increases the variance of the estimator. In general, there is a bias-variance trade-off and the bandwidth selection should depend on how slowly varying the underlying process for the parameters is, with larger bandwidths delivering smoother estimates. Giraitis et al. (2014) find that $H \propto T^{0.5}$ is asymptotically optimal (in the sense that it minimises the rate of the MSE of the estimator) but in DGP IV, we find that the estimator with slightly larger bandwidth $H = T^{0.55}$ delivers best RMSE performance, while the estimator with $H = T^{0.45}$ is best for coverage rates.

To summarise, in this section we demonstrate that the class of QBLL estimators exhibits good finite sample properties, comparable and in some cases superior to those of state space models under correct specification, and vastly outperforms competing estimators under misspecification of the state equation. The coverage rates of state space models are very sensitive to the priors and starting values and may suffer from non-decreasing RMSEs in case of misspecification of the

21

state equation's parametric assumptions. We find that inference based on the quasi-posterior distributions introduced in Section 3 not only delivers valid confidence intervals in all cases examined but is also robust to different processes for the drifting parameters.

5 Applications

5.1 Empirical application to U.S. inflation persistence

A question that has received attention and that divides the literature is whether the high inflation of the 1970s was an outcome of bad policy, suggesting the presence of time variation in the policy parameters of the model, or simply bad luck: a structural change in the size of shocks. In other words, was the subsequent period, referred to as the Great Moderation, a consequence of good policy under the Fed chairmanship of Paul Volcker and Alan Greenspan or was it a good luck event - a decrease in the volatilities of economic shocks. Supporters of the former view include Taylor (1993), DeLong (1997), Clarida, Gali and Gertler (2000) and Cogley and Sargent (2002, 2005). On the other hand, Sims (1980), Bernanke and Mihov (1998), Kim and Nelson (1999), McConnell and Perez Quiros (2000), Sims and Zha (2006) and Primiceri (2005) have insisted on explaining this phenomenon by the presence of changing volatility. A related question is whether inflation dynamics has changed after the Great Moderation period and, in particular, has inflation persistence decreased or remained constant over time as a consequence of the change in policy.

In this context, this section contributes to the literature on changing policy and inflation dynamics in the U.S. by applying the QBLL methodology proposed in this paper and revisiting classic results presented in Cogley and Sargent (2002, 2005), Primiceri (2005) and Cogley et al. (2010). Empirical contributions arise due to the novel approach of the paper which maintains an agnostic position on the parameter variation and, in turn, provides drifts in the parameters which are of nonparametric form and hence robust to state equation misspecification. Additional contributions arise from the longer sample period, which spans from 1954Q3-2015Q3.

We estimate a TVP BVAR(2) model using the macroeconomic series in Primiceri (2005) employing the QBLL approach introduced in Section 3. In addition, following Del Negro (2003)'s recommendation to include commodity prices due to their key role during the 1970s, the model is estimated with the addition of commodity prices to the set of variables. Details on the series, model specification, priors, robustness checks and some additional results can be found in Section 3.2 of the Supplementary Appendix. First, we re-write the VAR model in equation (8) in companion

form: $z_t = \mu_t + A_t z_{t-1} + \nu_t, \ \nu_t \sim \mathcal{N}(0, \Omega_t)$.

Figure 3: Core inflation and natural rate of unemployment
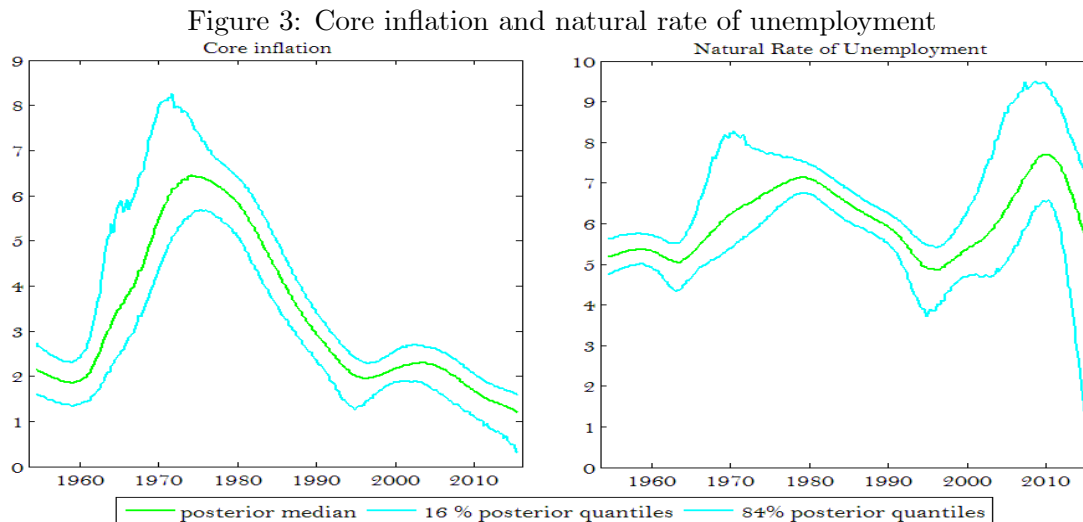
Figure 3 presents core inflation and the natural rate of unemployment, which are computed as the infinite horizon forecasts of inflation and unemployment implied by the model: $\tau_t = \lim_{h \to \infty} \mathbb{E}_t z_{t+h} \approx (I - A_t)^{-1} \mu_t$, where τ_t are the resulting long run means of z_t. The plots in Figure 3 are remarkably similar both quantitatively and in terms of shape to those presented in Cogley and Sargent (2002, 2005), even though their estimation method is considerably different. In particular, core inflation is around 2% in the 1960s, rises to 6.5% in the 1970s and then falls down to pre-1970s levels after mid-1980s. The natural rate of unemployment also peaks in the 1970s but much less sharply, which is also consistent with Cogley and Sargent (2002, 2005). More interestingly, since our sample contains the 2008 financial crisis, Figure 3 allows to assess the natural rate of unemployment during that period and it is evident that the increase is sharp and large in magnitude with the natural rate reaching 7.5% in 2009, when actual unemployment in the U.S. reached 10%. Figure 3 also suggests a recovery in the unemployment's natural rate after 2009. However, it is clear that the posterior confidence bands are much wider at the end of the sample implying larger uncertainty about its exact value after 2009. Figure 4 presents the conditional volatilities of the variables in the model over time. These are consistent with the results presented in Primiceri (2005). Inflation volatility peaks in the mid-1970s and then falls dramatically during the Great Moderation period. Unemployment volatility is high in the late 1950s (note that most of the earlier papers do not use data during the period 1950-1960, which is typically employed as a pre-sample). Moreover, unemployment volatility peaks in the mid-1970s and is relatively low during the 1990s. We find that the recent financial crisis causes not only a large increase in the unconditional mean of unemployment but also in its variance,

Figure 4: Conditional volatilities over time
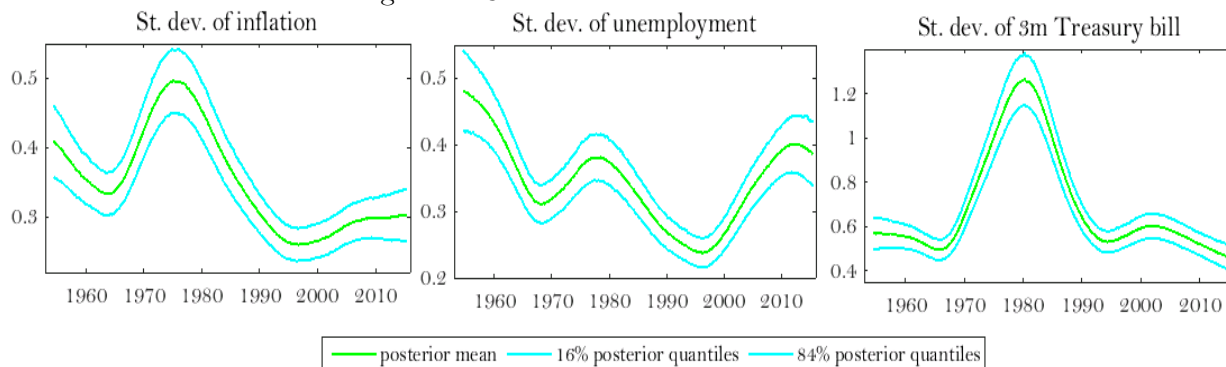
which remains high even after 2014. Finally, from Figure 4 we see that the 3-month Treasury rate's volatility exhibits a large increase in 1980 with aggressive monetary policy and tackling inflation under the leadership of Paul Volcker as chairman of the Fed and later falls to pre-1970s levels. Next, we turn to the question of inflation persistence implied by the QBLL approach. Cogley and

Figure 5: Inflation Persistence over time
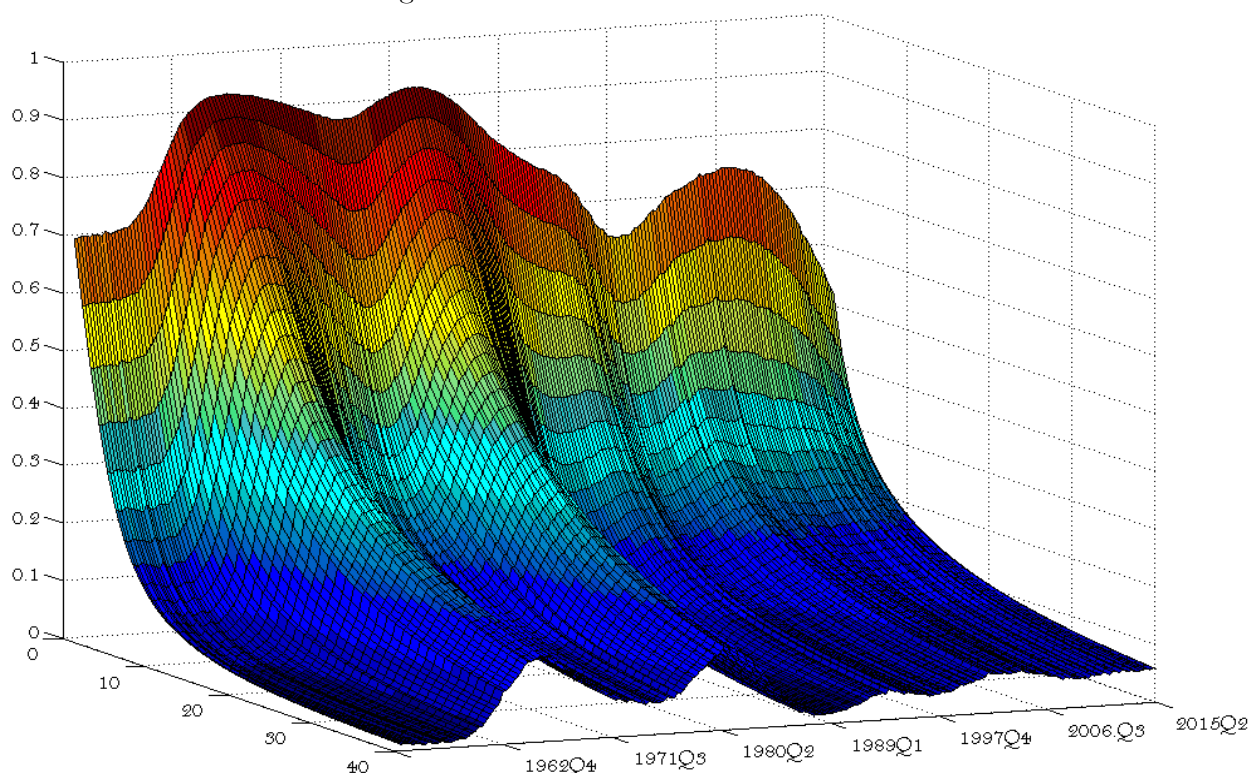
Sargent (2002) measure persistence as the normalised spectrum of inflation at zero. Cogley et al. (2010) introduce a new measure which they claim can be estimated more precisely. We follow their definition of inflation persistence h steps ahead as $R_{t,h}^2 = 1 - \frac{e_\pi \left[\sum_{j=0}^{h-1} A_t^j \Omega_t A_t^{j\prime}\right] e_\pi'}{e_\pi \left[\sum_{j=0}^{\infty} A_t^j \Omega_t A_t^{j\prime}\right] e_\pi'}$. The measure $R_{t,h}^2$

represents the proportion of total variation explained by past shocks (or equivalently one minus the proportion of total variation due to future shocks). It takes values between zero and unity, with large values implying that past shocks die out slowly making inflation more persistent and hence predictable. Figure 5 presents the inflation persistence, measured by the posterior mean of $R^2_{t,h}$, computed at each point in time and for horizons ranging from one quarter to ten years ahead. It is evident that inflation persistence dies out when the forecast horizon increases, which is expected. Furthermore, in the 1970s and early 1980s, inflation is very persistent with two peaks roughly corresponding to the two oil shocks and its persistence dies out slower over the forecast horizons. Furthermore, due to the longer sample, we can assess inflation persistence during the 2008 financial crisis, when quarterly inflation reached negative values in 2009 and it is clear from Figure 5 that inflation persistence increases during this period. However, unlike the period of 1970-1980s, in 2008 the $R^2_{t,h}$ measure dies out quickly over the forecast horizon and at 12 quarters ahead there is no difference in $R^2_{t,h}$ during the crisis compared to pre-crisis periods.

To summarise, we applied our novel QBLL approach to revisit much debated issues of changing macroeconomic dynamics, introduced in Clarida et al. (2000), Cogley and Sargent (2002, 2005), Primiceri (2005) and Cogley et al. (2010). We found that not only does the autoregressive component of the VAR change over time but also the volatility of the series, implying that whereas there has been a change in policy, especially after Paul Volcker's appointment as chairman of the Federal Reserve, there has also been a 'luck' component implied by the considerable time variation uncovered in the volatility of the series. Our QBLL estimator delivers results broadly consistent with evidence presented in Cogley and Sargent (2002, 2005) and Cogley et al. (2010), even though the estimation procedures differ considerably. In particular, unlike the parametric state space approach used in previous papers, the QBLL procedure is nonparametric with respect to the parameter processes and hence robust to various specifications. It is worth highlighting that the proposed methodology has the added advantage of being considerably simpler and computationally cheaper to implement and permits increasing the dimension of the VAR model to include additional relevant variables.

5.2 Forecasting exercise

In this section, we design a pseudo out-of-sample forecasting exercise in order to assess the forecasting record of the proposed QBLL approach. In particular, we compare the forecasts of various sizes BVAR models, with and without time varying parameters, using U.S. data. The dataset is from Stock and Watson (1996), containing 87 quarterly macroeconomic series and spanning from

1950Q2 to 2014Q1. Detailed variable descriptions and data transformations, as well as forecast origins, lag length and priors used can be found in Section 3.3 of the Supplementary Appendix.

We begin by estimating a set of small BVAR models which include 3 macroeconomic variables: GDP growth, inflation and the 3-month Treasury rate. The models we evaluate are: a fixed parameter BVAR model (F-BVAR), which we use as a benchmark, a time varying parameter and volatility BVAR model estimated using the closed form QBLL expressions derived in Proposition 2 (QBLL-TV), a drifting parameter homoscedastic BVAR model (QBLL-TVP) and an invariant parameter heteroscedastic BVAR (QBLL-TVV), estimated with the Gibbs algorithms proposed in Section 3.3 and outlined in Section 2.1 and 2.2 of the Supplementary Appendix respectively.

	RMSFEs small models			Forecast bias small models		
horizon	GDP growth	Inflation	T-bill	GDP growth	Inflation	T-bill
	F-BVAR					
1	0.82	0.35	0.20	0.15*	-0.05*	0.01
2	0.84	0.46	0.21	0.18*	-0.05	0.01
4	0.85	0.56	0.20	0.18*	-0.03	0.01
8	0.84	0.64	0.20	0.19*	0.02	0.01
	SS-TV					
1	1.01*	0.82*	1.00	0.16*	-0.03	0.01
2	1.00	0.80*	1.01	0.20*	-0.05	0.02
4	1.01	0.81*	1.00	0.21*	-0.08	0.02
8	1.01*	0.87*	1.05	0.23*	-0.08	0.02
	QBLL-TV					
1	1.03	0.94	1.05*	0.03	0.03	0.00
2	1.03	0.85*	1.05*	0.03	0.05	0.00
4	1.05	0.79*	1.04	0.02	0.06	0.00
8	1.04	0.77	1.04*	0.04	0.08	0.00
	SS-TVV					
1	1.02	0.82*	1.00	0.15*	-0.03	0.01
2	1.00	0.80*	1.02	0.20*	-0.04	0.02
4	1.01*	0.80*	0.99	0.21*	-0.05	0.02
8	1.01*	0.87*	1.00	0.21*	-0.04	0.02
	QBLL-TVV					
1	0.98	0.87*	1.00	0.15*	-0.03	0.01
2	1.02	0.86*	0.99	0.19*	-0.05	0.01
4	1.02	0.89*	1.01	0.20*	-0.07	0.01
8	1.02*	0.97	1.00	0.20*	-0.06	0.01
	SS-TVP					
1	1.02	0.83*	1.01	0.16*	0.00	0.02
2	1.00	0.81*	1.02*	0.20*	0.00	0.02
4	1.01	0.81*	0.99	0.20*	0.02	0.02
8	1.01	0.89*	1.00	0.20*	0.05	0.02
	QBLL-TVP					
1	1.04*	0.95	1.07*	0.03	0.03	0.00
2	1.04	0.85*	1.08*	0.02	0.05	0.00
4	1.05	0.78*	1.03	0.02	0.06	0.00
8	1.04	0.77	1.06*	0.03	0.08	0.00

Table 5. RMSFEs and forecast bias small models.

In addition, the use of a small dataset permits comparison with alternative state space TVP VAR models. For the comparison, we estimate three different state space models: one with both autoregressive parameters and volatilities varying over time, one with invariant volatilities and drifting parameters and one with invariant parameters and drifting volatilities, labelled SS-TV, SS-TVP

and SS-TVV respectively. Following the literature (Cogley and Sargent (2002), Primiceri (2005), Clark (2012), Mumtaz and Surico (2009)), the parameters, the elements of the lower triangular matrix and the log volatilities are all assumed to follow random walk processes. The left panel of Table 5 presents the absolute performance of the small time invariant F-BVAR model (in RMS-FEs) and the relative performance of alternative models over different horizons (numbers smaller than one imply superior performance of the particular model relative to F-BVAR). It is evident that the performance of all small time varying parameters models is very similar. In particular, they can rarely improve over the fixed parameter BVAR model for GDP growth and the 3-month Treasury rate, while all models deliver large and significant improvements in terms of RMSFEs for inflation. This is not surprising, given the evidence of considerable structural change in inflation dynamics presented in Section 5.1. The right panel of Table 5 presents the average forecast bias of the different models (positive numbers imply positive bias) and one, two and three stars indicate rejection of the null of zero bias at significance levels of 10%, 5% and 1% respectively. From the right panel of Table 5, the fixed parameter BVAR displays a large significant positive bias for GDP growth. This is expected as the evaluation periods span from 1970Q2 up to 2014Q2 and arguably contain periods characterised by serious structural change. To illustrate the issue, in the beginning of the Great Moderation period, for instance, the fixed parameter model delivers estimates by averaging the data from the oil crises insample periods and so its forecasts systematically underestimate output. On the other hand, TVP VAR models have a mechanism to capture the structural change both in the intercepts and in the autoregressive relationships between variables and hence reduce bias. Surprisingly, state space models featuring time variation in the parameters, and hence a mechanism to account for structural change, cannot improve systematic errors in the forecasts for GDP growth, implied by the statistically significant forecast bias. The gains of the QBLL-TV over the state space models could be explained by recalling that unlike the latter, the former approach does not force a parametric model on the parameter processes, which are allowed to vary freely and hence fit the data better. Another interesting result is that while allowing the variances to change with the QBLL-TVV model does not help with the bias, allowing for variation in the autoregressive parameters with the QBLL-TVP model makes GDP growth forecasts unbiased. This is an indication that variation in the autoregressive component of the VAR is important for reducing forecast bias. Table 6 accesses the quality of the density forecasts measured by log predictive score (LPS) and probability integral transformation (PIT) of the predictive density. The left panel of Table 6 displays absolute LPS for the F-BVAR model and differences in LPS for the

alternative models, so positive numbers imply superior performance over F-BVAR. The right panel
of Table 6 presents the p-values of the chi-squared Berkowitz (2001) test of uniformity of the PITs
for all models (numbers smaller than 0.05 indicate rejection of the null of uniform PITs at 5%). All
models can deliver statistically significant improvements for inflation over the benchmark in terms
of LPS. This is most likely a consequence of the superior point forecast implying that the forecast
density is centered more precisely around the ex-post realised value. In addition, some models
can also deliver superior forecast density performance for output growth and the interest rate. It
is evident that the models with drifting volatility and fixed parameters SS-TVV and QBLL-TVV
perform better than their counterparts SS-TVP and QBLL-TVP. This can be further seen from
the PIT p-values in Table 6, reinforcing the conclusion that for improving density forecasts, it is
crucial to allow for changing volatility rather than variation in the autoregressive parameters.

	Log score small models			PITs small models		
horizon	GDP growth	Inflation	T-bill	GDP growth	Inflation	T-bill
			F-BVAR			
1	-1.24	-0.42	0.10	0.00	0.00	0.00
2	-1.28	-0.67	0.01	0.00	0.00	0.00
4	-1.28	-0.87	0.03	0.00	0.13	0.00
8	-1.29	-1.04	-0.02	0.00	0.64	0.00
			SS-TV			
1	0.11***	0.33***	0.56***	0.02	0.73	0.23
2	0.08	0.42***	0.45***	0.00	0.02	0.54
4	0.03	0.44***	0.37**	0.00	0.00	0.40
8	0.00	0.37***	0.46*	0.00	0.01	0.02
			QBLL-TV			
1	0.02	0.22***	0.23**	0.03	0.43	0.00
2	0.01	0.32***	0.18	0.01	0.04	0.00
4	-0.01	0.36***	0.14	0.00	0.00	0.00
8	-0.07	0.24	0.18	0.00	0.00	0.00
			SS-TVV			
1	0.10***	0.33***	0.54***	0.03	0.95	0.45
2	0.10**	0.41***	0.44***	0.00	0.19	0.40
4	0.04	0.39***	0.36**	0.00	0.20	0.42
8	0.03	0.32***	0.47*	0.00	0.61	0.07
			QBLL-TVV			
1	0.06	0.31***	0.33***	0.00	0.07	0.00
2	0.04	0.36***	0.27**	0.00	0.45	0.00
4	0.03	0.34***	0.09	0.00	0.13	0.00
8	0.02	0.20**	0.23	0.00	0.00	0.00
			SS-TVP			
1	0.01	0.24***	0.05	0.01	0.79	0.00
2	0.03***	0.25***	0.08	0.00	0.24	0.00
4	0.01	0.21***	0.08	0.00	0.39	0.00
8	0.02	0.03	0.06	0.00	0.00	0.00
			QBLL-TVP			
1	-0.04**	0.10***	-0.08	0.00	0.00	0.00
2	-0.01	0.24***	-0.08	0.01	0.01	0.00
4	-0.02	0.34***	0.00	0.01	0.18	0.00
8	0.00	0.39**	0.05	0.00	0.03	0.00

Table 6. Log predictive scores and PITs small models.

Next, we investigate whether including additional variables to the model can improve the fore-
cast performance. This choice is motivated by the forecasting literature (Bańbura, Giannone and

Reichlin (2010), Koop (2011), Carriero, Clark and Marcellino (2015)), which suggests good fore-casting record of medium and large dimensional BVAR models. In particular, we estimate: (i) a fixed parameter model (F-BVAR), and (ii) a time varying parameter model with changing volatil-ity (QBLL-TV), on a medium size dataset (consisting of 17 macroeconomic variables described in Section 3 of the Supplementary Appendix). Using the medium dataset, we also estimate two mixtures of time varying and time invariant parameter models using Algorithms 1 and 2 developed in Sections 2.1 and 2.2 of the Supplementary Appendix respectively: QBLL-TVV and QBLL-TVP. Finally, we estimate two large BVAR models using the entire dataset (consisting of 87 variables): (i) a model with fixed parameters (F-BVAR-L), and (ii) a model, estimated with the QBLL approach, featuring time varying parameters and volatility (TV-QBLL-L).

horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
					RMSFEs medium and large models					
					F-BVAR					
1	0.77	0.34	0.40	0.03	0.73	0.07	1.25	0.20	0.85	0.50
2	0.82	0.41	0.53	0.04	0.71	0.08	1.60	0.20	0.86	0.65
4	0.84	0.50	0.59	0.04	0.68	0.09	1.57	0.20	0.88	0.72
8	0.84	0.62	0.70	0.04	0.69	0.09	1.59	0.20	0.88	0.85
					SS-TV					
1	1.08*	0.85*	-	-	-	-	-	1.00	-	-
2	1.03	0.89*	-	-	-	-	-	1.02	-	-
4	1.01	0.91*	-	-	-	-	-	1.00	-	-
8	1.02	0.90*	-	-	-	-	-	1.05	-	-
					QBLL-TV					
1	1.04	0.82*	0.87*	1.10*	0.86*	1.05	1.01	1.07	1.05	0.93
2	1.02	0.84*	0.88*	1.01	0.96	1.07	1.01	1.11*	0.99	0.95
4	1.05	0.79*	0.85*	1.08*	1.06	1.11*	1.10*	1.03	0.99	0.93
8	1.05	0.78	0.83	1.07*	1.08*	1.14*	1.08*	1.05*	1.01	0.90
					QBLL-TVV					
1	1.00	0.86*	0.97	1.14*	0.93	1.10	0.96	1.06	0.99	0.93*
2	1.01	0.84*	0.92*	0.99	0.93	1.03	0.99	0.97	0.99	0.92*
4	0.96*	0.81*	0.88*	0.98	0.97	1.02	0.99	1.00	1.01	0.91*
8	0.99	0.93*	0.96	1.01	1.01	1.00	0.97*	0.99	1.02	0.98
					QBLL-TVP					
1	1.11*	0.99	0.94	1.26*	0.85*	1.20*	1.16*	1.26*	1.01	1.01
2	1.05	0.93	0.88*	1.13	0.95	1.13*	1.05	1.18*	0.96	0.98
4	1.06	0.87	0.91	1.03	1.03	1.10*	1.04	1.14	0.99	0.95
8	1.07*	0.80	0.85	1.27*	1.01	1.32*	1.07*	1.27*	0.98	0.90
					F-BVAR-L					
1	0.96	0.94	0.99	0.99	0.90*	0.91	0.96	1.00	1.03	1.00
2	0.92*	0.93*	0.98	0.91*	0.93*	0.95	0.90	0.97	0.99	1.00
4	1.01	0.96	0.98	1.01	0.98	1.03	1.02	1.00	1.00	1.00
8	0.99	0.95*	0.96*	0.99*	1.00	1.01	0.99	1.01*	1.00	0.97
					QBLL-TV-L					
1	1.01	0.83*	0.92	1.14*	0.85*	0.96	0.99	1.07	1.09*	0.99
2	1.00	0.81*	0.89*	1.02	0.93	1.02	0.99	1.09*	0.99	0.96
4	1.08*	0.81*	0.87*	1.11*	1.07	1.14*	1.14*	1.04	1.00	0.95
8	1.08*	0.79	0.84	1.09*	1.09*	1.21*	1.13*	1.06*	1.02	0.90

Table 7. RMSFEs medium and large models.

Table 7 presents the absolute performance of the medium F-BVAR model (in RMSFEs) and the relative performance of alternative models over different horizons. From Table 7, it is clear that even a time invariant model with additional variables can outperform the small SS-TV model for

GDP growth. In addition, the various time varying medium and large models can outperform the SS-TV for inflation implied by the larger improvements over the fixed parameter model. The best performing model for inflation is the medium QBLL-TV model. Moreover, the time varying specifications can improve RMSFE performance of the F-BVAR for consumption and PCE growth.

horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
				Forecast bias medium and large models						
					F-BVAR					
1	0.14*	-0.01	-0.05	0.00	0.07	0.00	0.24*	0.01	0.17*	-0.08*
2	0.16*	-0.02	-0.06	0.00	0.11*	0.00	0.27*	0.01	0.21*	-0.11*
4	0.17*	-0.03	-0.07	0.00	0.13*	0.00	0.28	0.01	0.22*	-0.13
8	0.19*	0.00	-0.05	0.00	0.16*	0.00	0.33*	0.01	0.22*	-0.11
					SS-TV					
1	0.16*	-0.03	-	-	-	-	-	0.01	-	-
2	0.20*	-0.05	-	-	-	-	-	0.02	-	-
4	0.21*	-0.08	-	-	-	-	-	0.02	-	-
8	0.23*	-0.08	-	-	-	-	-	0.02	-	-
					QBLL-TV					
1	0.05	0.02	0.03	0.00	0.06	-0.01	0.07	0.00	0.03	0.03
2	0.02	0.04	0.04	0.00	0.05	0.00	0.03	0.00	0.01	0.04
4	0.01	0.05	0.04	0.00	0.04	0.00	-0.04	0.00	0.02	0.05
8	0.03	0.07	0.06	0.00	0.05	0.00	-0.01	0.00	0.03	0.06
					QBLL-TVV					
1	0.10*	0.00	-0.06*	0.00	0.12*	0.01	0.19*	0.01	0.17*	-0.06*
2	0.10	-0.01	-0.07*	0.00	0.13*	0.00	0.17	0.00	0.22*	-0.10*
4	0.10	-0.04	-0.09	0.00	0.15*	0.01	0.16	0.00	0.23*	-0.12
8	0.12	-0.05	-0.11	0.00	0.18*	0.01	0.20	0.00	0.24*	-0.15
					QBLL-TVP					
1	0.03	0.01	0.02	0.00	0.07	0.00	0.02	0.01	0.04	0.02
2	0.04	0.03	0.04	0.00	0.05	0.00	0.08	0.01	0.02	0.05
4	0.03	0.05	0.05	0.00	0.04	0.00	-0.01	0.00	0.03	0.06
8	0.03	0.07	0.07	0.00	0.07	0.01	0.01	0.00	0.03	0.06
					F-BVAR-L					
1	0.01	-0.06*	-0.07*	0.00	0.02	0.01*	0.01	0.00	0.08	-0.11*
2	0.07	-0.08*	-0.12*	0.00	0.08	0.01	0.10	-0.01	0.15*	-0.18*
4	0.16*	-0.11	-0.15*	0.00	0.14*	0.00	0.26	0.00	0.21*	-0.22*
8	0.20*	-0.08	-0.12	0.00	0.18*	0.00	0.36*	0.01	0.22*	-0.19
					QBLL-TV-L					
1	0.02	0.02	0.02	0.00	0.04	0.00	0.02	0.00	0.02	0.01
2	0.01	0.04	0.03	0.00	0.03	0.00	-0.01	0.00	0.02	0.03
4	-0.01	0.05	0.04	0.00	0.02	0.01	-0.09	0.00	0.03	0.05
8	0.00	0.07	0.05	0.00	0.04	0.01	-0.07	-0.01	0.03	0.06

Table 8. Forecast bias medium and large models.

Table 8 presents the forecast bias of the different models. It is evident that, while statistically significant forecast bias is present in medium and large time invariant models, the forecasts of the medium and large time varying specifications are virtually unbiased for all variables and all horizons. More interestingly, when we consider the medium model with drifting volatility and time invariant autoregressive component, we also find forecast bias in some variables. On the other hand, allowing for drifting autoregressive parameters when the volatility is held constant over time removes forecast bias. This confirms the results from the small models: variation in the autoregressive component of the VAR can significantly reduce systematic errors in the forecasts.

Table 9 accesses the quality of the density forecasts of the medium and large models. It displays

absolute LPS for the F-BVAR model and differences in LPS for the alternative models. We see from Table 9 that the various time varying specifications can deliver forecast density improvements for most variables. In particular, the best performing model is the large time varying parameter and volatility model, QBLL-TV-L, implying that by including additional variables we can obtain superior estimates of the uncertainty around the point forecast.

Finally, Table 10 presents the p-values of the chi-squared Berkowitz (2001) test of uniformity for all medium and large models. From Table 10, it is clear that for the time invariant models, both medium and large, we can reject the null of uniform PITs for most variables and horizons. On the other hand, for the medium time varying model, QBLL-TV, we cannot reject the null of uniformity for all variables one step ahead. The large time varying model, QBLL-TV-L, also performs exceptionally well. Another important result is that the QBLL-TVV model performs better in terms of PITs than the QBLL-TVP model, confirming our previous result that for improving density forecast performance, drifting volatility is more relevant than drifting autoregressive parameters.

horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
					Log score medium and large models					
					F-BVAR					
1	-1.19	-0.35	-0.51	2.03	-1.08	1.30	-1.67	0.08	-1.28	-0.70
2	-1.25	-0.56	-0.79	1.77	-1.13	1.08	-1.93	-0.02	-1.37	-0.96
4	-1.28	-0.79	-0.94	1.76	-1.13	0.99	-1.96	-0.03	-1.37	-1.16
8	-1.29	-0.99	-1.23	1.75	-1.14	0.98	-1.97	-0.02	-1.37	-1.35
					SS-TV					
1	0.05	0.26***	-	-	-	-	-	0.59***	-	-
2	0.04	0.31***	-	-	-	-	-	0.48***	-	-
4	0.03	0.35***	-	-	-	-	-	0.43**	-	-
8	0.00	0.33***	-	-	-	-	-	0.45*	-	-
					QBLL-TV					
1	0.07	0.29***	0.19**	-0.03	0.19***	-0.01	0.12**	0.29**	0.08	0.13
2	0.02	0.35***	0.15**	0.00	0.15***	-0.11	0.04	0.24*	0.13	0.07
4	0.03	0.39***	0.09	-0.04	0.05	-0.22	-0.11	0.22	0.11	0.11
8	0.00	0.33**	0.07	-0.11	-0.04	-0.34	-0.08	0.22	0.10	-0.01
					QBLL-TVV					
1	0.06	0.20***	-0.01	-0.08	0.09**	-0.08	0.17***	0.32**	0.06	0.04
2	0.07	0.27***	0.08**	0.01	0.08	-0.05	-0.04	0.34**	0.13	0.10**
4	0.12*	0.32***	0.18***	0.01	0.11*	-0.05	-0.01	0.22*	0.12	0.18**
8	0.07	0.12	0.03	-0.03	0.04	-0.03	0.06	0.22	0.09	-0.03
					QBLL-TVP					
1	-0.08	0.13***	0.07	-0.22***	0.07	-0.19***	-0.24***	-0.25**	-0.05	0.01
2	-0.05*	0.18***	0.08	-0.18***	0.06	-0.18***	-0.08*	0.04	0.04	-0.03
4	-0.03	0.25***	0.09	-0.23***	0.03	-0.21***	-0.03	0.09	0.04	0.13
8	-0.06**	0.32*	0.40*	-0.34***	-0.01	-0.33***	-0.02	-0.08	0.09	0.27*
					F-BVAR-L					
1	0.07	0.07***	-0.02	0.01	0.07**	0.07**	0.04**	0.07	-0.04	0.01
2	0.08***	0.08***	0.03	0.09***	0.05**	0.07*	0.10***	0.08	0.05	0.00
4	0.00	0.07***	0.05*	0.00	0.01	-0.01	0.00	0.08	0.04	0.02
8	0.01	0.05**	0.07**	0.03*	-0.01	-0.02	0.00	-0.06	0.01	0.03
					QBLL-TV-L					
1	0.09	0.30***	0.07	-0.06	0.20***	0.09*	0.16***	0.38***	0.04	-0.04
2	0.08*	0.37***	0.16**	0.05	0.17***	0.06	0.13*	0.31**	0.12	0.07
4	-0.02	0.39***	0.09	-0.04	0.05	-0.16	-0.07	0.22	0.11	0.12
8	-0.03	0.35**	0.09	-0.12	-0.04	-0.37	-0.11	0.16	0.10	0.06

Table 9. Log predictive scores medium and large models.

horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
					PITs medium and large models					
					F-BVAR					
1	0.04	0.00	0.09	0.00	0.00	0.11	0.00	0.00	0.00	0.00
2	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.13
8	0.00	0.75	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05
					SS-TV					
1	0.02	0.73	-	-	-	-	-	0.23	-	-
2	0.00	0.02	-	-	-	-	-	0.54	-	-
4	0.00	0.00	-	-	-	-	-	0.40	-	-
8	0.00	0.01	-	-	-	-	-	0.02	-	-
					QBLL-TV					
1	0.61	0.32	0.52	0.93	0.13	0.35	0.56	0.10	0.07	0.59
2	0.17	0.25	0.00	0.14	0.24	0.00	0.00	0.34	0.09	0.00
4	0.42	0.06	0.00	0.38	0.29	0.00	0.00	0.00	0.15	0.00
8	0.14	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.09	0.00
					QBLL-TVV					
1	0.10	0.04	0.00	0.01	0.01	0.00	0.06	0.00	0.00	0.00
2	0.15	0.48	0.01	0.27	0.02	0.00	0.00	0.68	0.00	0.00
4	0.12	0.03	0.20	0.08	0.00	0.00	0.00	0.00	0.00	0.03
8	0.08	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00
					QBLL-TVP					
1	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.03
2	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00
4	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.71	0.25
					F-BVAR-L					
1	0.05	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
8	0.00	0.04	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00
					QBLL-TV-L					
1	0.81	0.16	0.03	0.82	0.12	0.76	0.22	0.00	0.10	0.00
2	0.72	0.13	0.00	0.94	0.12	0.38	0.62	0.44	0.19	0.00
4	0.13	0.14	0.00	0.66	0.60	0.00	0.00	0.05	0.38	0.00
8	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.02	0.29	0.00

Table 10. PITs medium and large models: p-values of the Berkowitz (2001)'s test.

To summarise, a forecasting exercise was designed, where we estimated various sizes fixed and time varying parameter BVAR models. Several conclusions emerged: first, the small QBLL models can deliver forecast performance similar to that of state space models. Second, the QBLL approach allows to increase the number of variables in the system: as we demonstrated increasing the VAR dimension while allowing for time variation in the parameters can improve both point and density forecasts over (i) small models, and (ii) time invariant models of the same dimensions. Third, the QBLL approach performs exceptionally well in eliminating forecast bias as well as in delivering uniform PITS and this is true for most variables and horizons. Finally, by assessing the forecast performance of VAR models with mixtures of time varying and time invariant parameters, we reached an important conclusion: allowing for drifting volatility in a VAR model improves density forecasts compared to a model with invariant volatility; on the other hand, time variation in the intercept and autoregressive component of the VAR model can significantly reduce systematic forecast errors. Since both point and density performance is crucial for forecasting, we conclude that

variation in both the parameters and the covariance matrix of VAR models should be considered.

6   Conclusion

This paper establishes a novel quasi-Bayesian local likelihood (QBLL) approach for econometric inference in models with time varying parameters. The Bayesian framework is based on augmenting the local likelihood of Giraitis et al. (2016) with a prior distribution; this augmentation principle delivers asymptotically valid quasi-posterior distributions which admit closed form expressions in the special case of a Gaussian VAR model. The approach is of sufficient generality and flexibility to give rise to Gibbs algorithms that are able to sample from a VAR model with a mixture of time varying and time invariant parameters.

Our Monte Carlo exercise demonstrates that the class of QBLL estimators based on: (i) the closed form distribution derived in Section 3.1; (ii) the proposed MCMC algorithms in Section 3.3, both exhibit good finite sample properties, and inference based on their quasi-posterior densities delivers valid confidence intervals. Importantly, QBLL inference is robust to different processes for the drifting parameters, as its validity does not depend on parametric restrictions typically imposed by state space models.

The novel QBLL approach is employed to empirically address the issue of changing macro-economic dynamics in the U.S. and confirms previous results of Cogley and Sargent (2002, 2005) and Primiceri (2005) on the presence of significant structural change in core inflation, inflation persistence and the natural rate of unemployment as well as of substantial drifts in the volatility of the series. In the light of these results, we conclude that ignoring the time variation (by instead estimating a time invariant model) will result in invalid inference on the model's parameters.

Finally, in a forecasting exercise we find that the QBLL estimators deliver excellent forecast performance. Their ability to accommodate the drifting parameters nonparametrically delivers forecast improvements over fixed parameter models, such as virtually unbiased forecasts and uniform PITs. In addition, their capacity to considerably increase the number of variables can improve forecast performance (both point and density forecasts) of small-dimensional models. We find that both drifts in the volatility and in the parameters can have important impact on forecasting: drifts in the volatility improve density forecasts while time variation in the autoregressive component significantly reduces forecast bias.

References

Abramovich, Y., Spencer, N. and Turley, M. (2007). Time-varying autoregressive (TVAR) models for multiple radar observations, *IEEE Transactions on Signal Processing* **55**(4): 1298–1311.

Bańbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions, *Journal of Applied Econometrics* **25**(1): 71–92.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics* **19**(4): 465–474.

Bernanke, B. and Mihov, I. (1998). Measuring monetary policy, *Quarterly Journal of Economics* **113**: 869–902.

Canova, F. and Perez Forero, F. (2015). Estimating overidentified, nonrecursive, time-varying coeffcients structural vector autoregressions, *Quantitative Economics* **6**(2): 359–384.

Carriero, A., Clark, T. and Marcellino, M. (2015). Bayesian VARs: specification, choices and forecast accuracy, *Journal of Applied Econometrics* **30**: 46–73.

Carriero, A., Clark, T. and Marcellino, M. (2016). Large vector autoregressions with stochastic volatility and flexible priors, *Working Paper* .

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation, *Journal of Econometrics* **115**: 239–346.

Clarida, R., Gali, J. and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory, *Quarterly Journal of Economics* **115**(1): 147–180.

Clark, T. E. (2012). Real-time density forecasting from BVARs with stochastic volatility, *Journal of Business and Economic Statistics* **29**(3): 327–341.

Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010). Inflation-gap persistence in the US, *American Economic Journal: Macroeconomics* **2(1)**: 43–69.

Cogley, T. and Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics, *in* B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 331–88.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post World War II US, *Review of Economic Dynamics* **8**: 262–302.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes, *The Annals of Statistics* **28**(6): 1762–1794.

Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes, *The Annals of Statistics* **34**(6): 2790–2824.

Del Negro, M. (2003). Discussion of Cogley and Sargent's 'drifts and volatilities: Monetary policy and outcomes of the post WWII U.S., *Federal Reserve of Atlanta Working Paper 26* .

DeLong, J. B. (1997). Reducing inflation, *in* C. Romer and D. Romer (eds), *America's only peacetime inflation: the 1970's*, NBER Studies in Business Cycles.

Giraitis, L., Kapetanios, G. and Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change, *Journal of Econometrics* **177**: 153–170.

Giraitis, L., Kapetanios, G., Wetherilt, A. and Zikes, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market, *Journal of Applied Econometrics* **31**(1): 58–84.

Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models, *Journal of Econometrics* **179**(1): 46–65.

Giraitis, L., Kapetanios, G. and Yates, T. (2018). Inference on multivariate heteroscedastic stochastic time varying coefficient models, *Journal of Time Series Analysis* **39**(2): 129–149.

Jacquier, E., Polson, N. G. and Rossi, P. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**: 371–418.

Kapetanios, G., Marcellino, M. and Venditti, F. (2015). Large time-varying parameter VAR: kernel-based mixed estimators, *Working Paper* .

Kapetanios, G. and Zikes, F. (2017). Time varying lasso, *Working Paper* .

Kim, C. and Nelson, C. R. (1999). Has the U.S. Economy become more stable? A Bayesian Approach based on a Markov-Switching Model of the Business Cycle, *Review of Economics and Statistics* **81**(4): 608–618.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**: 361–393.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**(1): 1 – 25.

Koop, G. (2011). Forecasting with medium and large Bayesian VARs, *Journal of Applied Econometrics* .

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs, *Journal of Econometrics* **177**: 185–198.

McConnell, M. and Perez Quiros, G. (2000). Output fluctuations in the U.S.: what has changed since the early 1980s?, *American Economic Review* **90**: 1464–1476.

Mumtaz, H. and Surico, P. (2009). Time-varying yield curve dynamics and monetary policy, *Journal of Applied Econometrics* **24**(6): 895–913.

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

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

Sims, C. A. (2001). Comment on Sargent and Cogley's 'evolving post WWII US inflation dynamics, *NBER Macroeconomics Annual* **16**: 373–379.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 1193–1224.

Stock, J. (2001). Discussion of Sargent and Cogley's Evolving post WWII US inflation dynamics, *NBER Macroeconomics Annual* **16**: 379–387.

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

Taylor, J. B. (1993). Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* **39**: 195–214.

Tian, L., Liu, J. S. and Wei, L. (2007). Implementation of estimating function-based inference procedures with Markov chain Monte Carlo samplers, *Journal of American Statistical Associations* **102**(479): 881–888.