Towards Robust, Scalable and Interpretable Time Series Forecasting using Bayesian Vector Auto-Regression

Rishab Guha [1] **Eric Aldrich** [1] **Bertram Ieong** [1] **Shanshan Li** [1] **Domenico Giannone** [2] **Juan Huerta** [1]

Abstract

We present a flexible, scalable, and interpretable framework for automated forecasting of multivariate time-series, building off of the Bayesian Vector Autoregression (BVAR) literature in macroeconometrics. Our algorithm allows for full posterior estimates of hundreds of interaction parameters, with minimal hand-tuning or hyperparameter specification required. The model can be easily extended to account for non-stationary breaks such as the COVID-19 pandemic. In experiments our model outperforms comparably-flexible time-series models at forecasting inflation.

1. Introduction

Time-series forecasts can often be substantially improved by modeling multiple related time-series jointly. However, relative to univariate forecasting problems, there are few automated methods available for producing joint distributional forecasts of time-series data without substantial manual fine-tuning. Additionally, existing methods for multivariate forecasting, such as vector auto-regressions, produce forecasts which are black-box convolutions of the joint distribution of all modeled series and are often difficult to understand and analyze.

We propose a Bayesian approach to estimating a vector auto-regression (VAR) with time-varying means, and present prior specifications that can be used to make estimation and forecasting effectively automatic given a sample of appropriate length. Our model uses strong priors, which can be automatically, estimated from a pre-sample, to regularize coefficient estimation. The use of Bayesian estimation automates much of the parameter selection and fine-tuning that is necessary when estimating standard VARs. The time-varying mean component of our model allows for highly interpretable medium and long-term forecasts that can be optionally specified as a factor structure, while the VAR component allows for short-term reactions to high-frequency shocks.

Our model builds off of a variety of VAR models in the macroeconometric literature, which have been specialized for the analysis of specific macroeconomic scenarios, e.g., (Bańbura et al., 2010), (Del Negro et al., 2019) (Christiano et al., 1999). Relative to these models, in which the model structure is tightly connected to the statistical likelihood function, we present a fully general framework that can be immediately and automatically applied to a wide variety of time series. We generalize the estimation procedure for VARs by expressing our model in state-space form, and developing an estimation algorithm that nests the Kalman simulation smoother within a Gibbs Sampler to produce posterior estimates that can be easily parallelized and efficiently computed.

Most methods for multivariate forecasting require that the input data be transformed into stationary time-series before estimation. For many time-series encountered in applied forecasting applications, such as macroeconomic variables, or product demand data, a common approach is to take year-on-year growth rates or one-year log differences. However, these transformations fail when confronted with sudden level shifts such as those generated by the COVID-19 pandemic. Because such level shifts affect both the numerator and the denominator of a year-on-year growth calculation they can cause ”baseline effects” which make forecasting difficult in the year following a shift. We propose a simple extension of our model in which a (potentially time-varying) level shift is integrated into the state-space, so that the model can jointly estimate the effect of the shift on both the numerator and denominator of realized growth rates.

We apply our model to the task of forecasting core CPI, the most commonly-tracked measure of headline inflation, in the United States. We show that our model, which models the joint evolution of core CPI with its components outperforms a variety of off-the-shelf competitor forecasting methods in backtests, with minimal manual specification required. We additionally show that our adjustment for level shifts during the COVID-19 period is able to account

[1]Amazon, New York, New York, United States [2]Amazon, Seattle, Washington, United States. Correspondence to: Rishab Guha <rishguha@amazon.com>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

for baseline effects, and predict the realized year-over-year increase in core CPI that surprised many macroeconomic forecasters in April 2021.

2. The statistical model

Let y_t be an n-dimensional vector of observables which have been transformed into stationarity; in practice, we usually work with year-on-year growth rates or log-differences. Our basic estimation equation separates the observable data into a time-varying mean $\bar{y}_t$ and a VAR component, $\tilde{y}_t$:

$$y_t = \bar{y}_t + \tilde{y}_t \tag{1}$$

We allow the time-varying mean component to be driven by a k-dimensional unobserved factor structure, γ_t, with the factor loadings captured in an $n \times k$ dimensional matrix Λ, and the factors themselves evolving according to a random walk:

$$\bar{y}_t = \Lambda \gamma_t \tag{2}$$

$$\gamma_t = \gamma_{t-1} + u_t \tag{3}$$

with $u_t \sim \mathcal{N}(0, \Sigma_u)$. The factor specification is fully optional, and can be trivialized by setting $\Lambda = I_{n\times n}$. Setting Λ to be a loading matrix, estimated within the Gibbs sampler, with $k < n$, will allow for dimensionality reduction, whereas setting Λ to a set of deterministic loadings (e.g., 0s and 1s) with $k \geq n$ allows for an overidentified factor structure that can ease interpretation and prior specification.

$\tilde{y}_t$ captures the high-frequency fluctuation of each part of y_t around its low-frequency mean. These fluctuations will have an unconditional mean of zero, and evolve according to a VAR with lag length p. Specifically, for each $i = 1, \dots, n$ we will have

$$\tilde{y}_{i,t} = \left[\sum_{k=1}^{n} \sum_{s=1}^{p} \beta_{i,k,s} \cdot \tilde{y}_{k,t-s} \right] + \epsilon_{i,t} \tag{4}$$

where $(\epsilon_{1,t}, \dots, \epsilon_{n,t}) \equiv \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$

2.1. Estimating the model

For estimation, we express our model in an unobserved-components state-space framework. We then use the simulation smoother of (Carter & Kohn, 1994) or (Durbin & Koopman, 2002) to sample the trend and transitory components from their posterior distribution conditional on the observed data and hypothesized values for the model parameters, β, Σ_u, and Σ_ϵ. Then, given these samples, we can use standard Bayesian VAR estimation techniques to sample the parameters from their posterior given the states. Building off of the approach of (Bańbura et al., 2010) we use the Gibbs Sampler approach of (Chib & Greenberg, 1994) to generate a chain of draws from the full joint posterior.

To describe our algorithm, let $\gamma_t^{(j)}$ and $\tilde{y}_t^{(j)}$ be the j-th samples of the unobserved states, and let $\beta^{(j)}$, $\Sigma_u^{(j)}$, and $\Sigma_\epsilon^{(j)}$ be the j-th samples of the model parameters. Let ϕ_γ and ϕ_u be vectors collecting the parameters for the priors of the low-frequency factors' initial state and innovation variances respectively. To derive the $j+1$-th samples, we apply the following algorithm:

1. Apply the Kalman simulation smoother to the observed data, y_t, with the j-th draw parameters encoded in the system matrices, to produce a new draw of the trend and transitory components from their posterior:

$$\gamma_t^{(j+1)}, \tilde{y}_t^{(j+1)} \sim P(\gamma_t, \tilde{y}_t | y_t, \beta^{(j)}, \Sigma_u^{(j)}, \Sigma_\epsilon^{(j)}, \phi_\gamma) \tag{5}$$

2. Produce estimates of the random-walk innovation term for the trend, $u_t^{(j+1)}$, and use these estimates to sample from the posterior of the variance-covariance matrix of the trend innovations:

$$u_t^{(j+1)} \equiv \gamma_t^{(j+1)} - \gamma_{t-1}^{(j+1)} \tag{6}$$

$$\Sigma_u^{(j+1)} \sim P(\Sigma_u | u_t^{(j+1)}, \phi_u) \tag{7}$$

3. Sample the VAR parameters from their posterior. This requires two steps. First, we estimate a Bayesian VAR, with a Minnesota prior as described in (Litterman, 1986), to sample proposals of the VAR parameters from their posterior given the transitory components:

$$\hat{\beta}^{(j+1)}, \hat{\Sigma}_\epsilon^{(j+1)} \sim P(\beta, \Sigma | \tilde{y}_t) \tag{8}$$

These sampled parameters will reflect the likelihood of each observation $\tilde{y}_t$ conditional on its p predecessors $\tilde{y}_{t-1}, \dots, \tilde{y}_{t-p}$. However, they do not incorporate the likelihood that we observe the first p terms, $\tilde{y}_{-p+1}, \dots, \tilde{y}_0$ given the model's implied steady state.

To account for this, we first derive the steady state variance-covariance matrix of the transitory terms, $\hat{V}^{(j+1)}$ as the solution to the Lyapunov forward equation:

$$\hat{V}^{(j+1)} = C^{(j+1)} \cdot \hat{V}^{(j+1)} \cdot C^{(j+1)^T} + \Sigma_\epsilon^{(j+1)} \tag{9}$$

where $C^{(j+1)}$ is the transitory term's transition matrix which is defined as a function of $\beta^{(j+1)}$. We use the likelihood of the initial conditions, conditional on this steady-state covariance matrix to perform a Metropolis-Hastings step. This has the effect of rejecting parameter estimates which imply initial conditions that are very far away from steady-state.

2.2. Extension to non-stationary level shifts

In most applications the transformed stationary variables contained in y_t will be some form of year-on-year growth rate. This type of modeling normally works very well for a variety of economic series, but encounters problems when faced with sudden level shifts. The presence of such level shifts in both the numerator and the denominator of the year-on-year growth calculation causes non-stationarity.

For example, during the COVID-19 pandemic a variety of macroeconomic series saw sudden shocks to their year-on-year growth rates during the spring of 2020, offset by shocks of the opposite sign in spring 2021.The generality of our state-space setting allows us to extend our model to allow for such level shifts, given knowledge of their location.

Let x_t be the log-level of all variables under consideration, and let x_{t-l} be the one-year lag; for example, for monthly data we will have $l = 12$. Then we extend the state space model to allow for a time-varying level shift, S_t, in the factor structure, with an indicator variable ξ_t used to mark the periods in which the shift is active:

$$x_t = \Lambda \cdot S_t \cdot \xi_t + y_t + x_{t-l} \quad (10)$$

The "normal-times" growth component, y_t evolves exactly as specified in the previous section. In our baseline specification, level shift evolves according to a random walk without drift. In experiments, we have found that this restriction can be relaxed by sampling the autoregressive parameter from a discrete proposal distribution, and applying a Metropolis-Hastings step based on the likelihood of the full model. This will estimate a mixture model across discrete scenarios for the dynamics of the shift. A similar approach can be applied to estimating the size of the shocks to the level shift, relative to the size of the shocks to the baseline drift terms.

3. Empirical Application

We use our model to produce forecasts of the Consumer Price Index for All Urban Consumers ex food and energy, commonly denoted core CPI, which is one of the most commonly-tracked measures of macroeconomic inflation. To showcase the multi-variate capabilities of our method, we jointly model year-over-year changes in core CPI along with year-over-year changes in its seven primary components. We retrieve all data from the FRED macroeconomic database maintained by the Federal Reserve Bank of St. Louis. With 8 variables and 3 monthly lags, our model has 264 parameters to estimate with 15 years of data, highlighting the importance of strong priors for regularization.

Figure 1 shows the model-implied decomposition of year-over-year changes in CPI into persistent and transitory variation for Core CPI as well as the housing component of CPI from the model estimated through January of 2012. The persistent components of each series are clearly very high correlated, but the housing component shows stronger price growth during the boom years of 2006-2007 followed by price declines during the house crash that accompanied the Great Recession. The uncertainty attributable to persistent variation grows over time, following the random walk specification, whereas the uncertainty due to the transitory component stays of constant width.

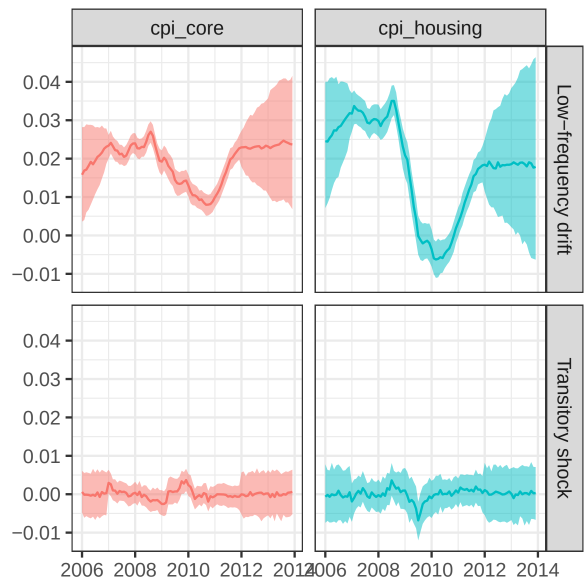

Figure 1. Model decomposition of year-over-year inflation in Core CPI and housing CPI from data through January of 2012, with forecast window through the end of 2013.

To compare our model's out-of-sample forecast performance with that of other "plug and play" models, we perform a horse-race backtest of Core CPI prediction against the Prophet model described in (Taylor & Letham, 2018), as well as a standard vector autoregression implemented using the `vars` package. Our proposed model is denoted BVAR-SSM. For each month between January of 2008 and January of 2019 we estimate each model on all available data, and forecast forward 12 months. As Table 1 shows, our proposed model shows substantially lower mean average percentage error than its competitors when compared across a variety of horizons. A key advantage of our proposed model is that it can match the short-horizon predictive performance of a "standard" vector auto-regression, without the long-horizon overreaction that tends to characterize such models. This is due to the BVAR-SSM model's slowly time-varying intercept, as well as the regularization that the Minnesota Prior applies to the cross-variable autocorrelation terms.

To test the performance of our extended state-space model

Table 1. Mean Average Percentage Error, in basis points, for proposed model vs. competitors on CPI backtest, by forecast horizon in months. The degradation in performance at 12-months is due to difficulties in forecasting year-on-year growth rates more than one year ahead.

HORIZON	BVAR-SSM	PROPHET	VAR
1	15.85	36.95	15.96
2	19.66	38.29	22.84
3	22.86	39.72	30.37
4	26.06	41.16	35.54
5	29.33	42.64	38.44
6	30.21	43.96	40.66
7	34.34	45.12	42.89
8	38.10	46.25	44.96
9	40.16	47.26	47.79
10	42.94	48.24	51.42
11	45.21	49.15	53.98
12	54.37	50.29	57.93

for non-stationary level shifts, as described in 2.4, we apply it to recent data through the COVID-19 pandemic period. We specify that the level shift begins in April of 2020, but give the model no additional guidance about its size or evolution. As Figure 2 shows, the level shift term captures most of the pandemic-driven variation in realized CPI inflation, allowing the low-frequency estimate of underlying inflation to remain stable. The shift also captures differential sectoral effects: while core CPI decreased slightly, prices for food and beverages actually increased.

The benefit of adjusting for non-stationary level shifts is apparent when comparing forecasts for April 2021. April 2020 marked the start of the COVID-19 pandemic's impact on the US economy, causing a sudden sharp decline in the consumer price index. This implies a predictable increase in year-over-year inflation 12 months later. As Figure 3 shows, the BVAR-SSM model with level shifts is able to account for these dynamics, whereas the Prophet and VAR models cannot.

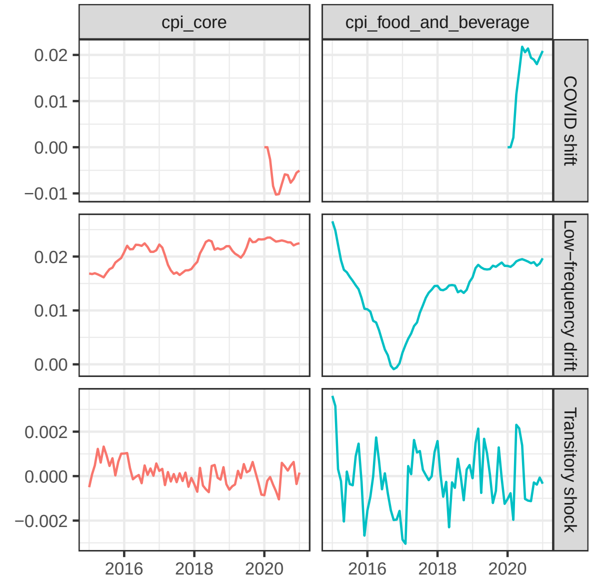

Figure 2. Model decomposition of year-over-year inflation in Core CPI and food/beverage CPI from data through 2020.

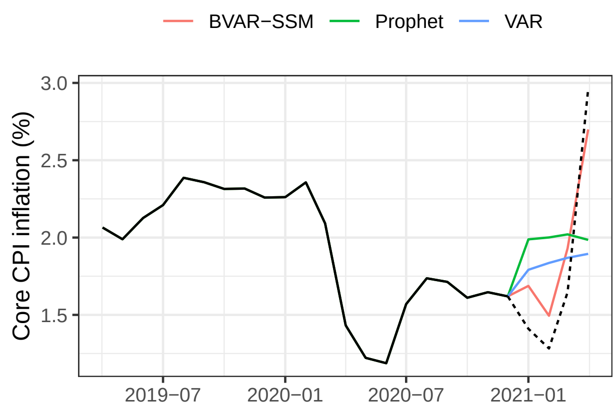

Figure 3. Forecast performance for models trained through December of 2020. Actuals are in black, and actuals after the training vintage are dashed.

4. Conclusion

We have presented a flexible statistical model for Bayesian multivariate time-series forecasting. Our model preserves the conceptual simplicity and interpretability of a low-frequency random-walk model, while allowing for complex cross-variable interactions through a high-frequency vector autoregression. We specify strong priors, derived from a large literature in applied macroeconometrics where they have proved suitable for a wide variety of time series, to discipline estimation and prevent overfitting. We show that our model can be easily extended to accommodate non-stationary level shifts, such as those seen during the COVID-19 pandemic, using state-space methods.

In an empirical application, we find that our model outperforms competitor models at the task of forecasting inflation. Our model generates interpretable insights into the multidimensional factor structure of low-frequency trends in inflation, while outperforming more black-box models at short horizons. With its state-space extension our model is able to account for the baseline effects caused by the COVID-19 pandemic.

References

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

Carter, C. K. and Kohn, R. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Chib, S. and Greenberg, E. Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183–206, 1994. URL https://EconPapers.repec.org/RePEc:eee:econom:v:64:y:1994:i:1-2:p:183-206.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. Chapter 2 monetary policy shocks: What have we learned and to what end? volume 1 of *Handbook of Macroeconomics*, pp. 65–148. Elsevier, 1999. doi: https://doi.org/10.1016/S1574-0048(99)01005-8. URL https://www.sciencedirect.com/science/article/pii/S1574004899010058.

Del Negro, M., Giannone, D., Giannoni, M. P., and Tambalotti, A. Global trends in interest rates. *Journal of International Economics*, 118:248 – 262, 2019. ISSN 0022-1996. doi: https://doi.org/10.1016/j.jinteco.2019.01.010. URL http://www.sciencedirect.com/science/article/pii/S0022199618302927.

Durbin, J. and Koopman, S. J. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616, 2002.

Durbin, J. and Koopman, S. J. *Time Series Analysis By State Space Methods*. Oxford University Press, 2012.

Litterman, R. B. Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38, 1986. ISSN 07350015. URL http://www.jstor.org/stable/1391384.

Taylor, S. J. and Letham, B. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018. doi: 10.1080/00031305.2017.1380080. URL https://doi.org/10.1080/00031305.2017.1380080.