IFC-Bank of Italy Workshop on "Machine learning in central banking"

19-22 October 2021, Rome / virtual event

Applications of variational inference in the Bank of Russia[1]

Sergei Seleznev and Ramis Khabibullin,
Bank of Russia

[1] This presentation was prepared for the Workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the central banks and other institutions represented at the event.

Bank of Russia

APPLICATIONS OF VARIATIONAL INFERENCE IN THE BANK OF RUSSIA

Ramis Khabibullin
Bank of Russia,* HSE

Sergei Seleznev
Bank of Russia*

19 October 2021
IFC and Bank of Italy Workshop ‘Data Science in Central Banking’
Part 1: Machine Learning Techniques

*The views expressed in the presentation are solely those of the authors and do not necessarily represent the official position of the Bank of Russia. The Bank of Russia is not responsible for the contents of the presentation.

Bayesian estimation of macro models

Vector Autoregressions

- Litterman (1980), Doan, Litterman and Sims (1984), Sims (1993), Villani (2009), Banbura, Giannone and Reichlin (2010), Koop and Korobilis (2010), Giannone, Lenza and Primiceri (2015).

Dynamic Factor Models

- Otrok and Whiteman (1998), Kim and Nelson (1998), Aguilar and West (2000), Blake and Mumtaz (2012).

Dynamic Stochastic General Equilibrium Models

- Smets and Wouters (2003, 2007), Fernandez-Villaverde and Rubio-Ramirez (2007), Justiniano and Primiceri (2008), Herbst and Schorfheide (2015).

Agent Based Models

- Grazzini, Richardi and Tsionas (2017), Gatti and Grazzini (2018), Lux (2018).

Bayesian techniques for intractable posterior

Sampling (asymptotically sample from exact posterior)

- Gibbs Sampling (Casella and Goerge (1992));
- Importance Sampling (Owen (2013));
- Metropolis-Hastings (Chib and Greenberg (1995));
- Hamiltonian Monte Carlo (Neal (2011));
- No-U-Turn Sampling (Hoffman and Gelman (2014));
- Sequential Monte Carlo (Doucet, De Freitas and Gordon (2001));
- PDMP Continuous-Time Monte Carlo (Fearnhead et al. (2016)).

Optimization (allows to estimate larger models)

- MAP estimation;
- Expectation Propagation algorithm (Minka (2001));
- Variational Bayes estimation (Wainwright and Jordan (2008));
- α-divergense (Li and Turner (2016)).

What does VB offer?

Approximate posterior

- Chooses the posterior approximation from a family of distributions that allows independent sampling.

Solving via optimization

- Finds an approximate posterior minimizing KL divergence between posterior and approximation family;
- Allows achieving the same accuracy faster than MCMC methods;
- Allows estimating posterior simultaneously maximizing the marginal likelihood with respect to hyperparameters.

Large-scale applications

- Estimates models with thousands/dozens of thousands of parameters on Desktop PC.

Variational Bayes

A VB algorithm maximizes the lower bound of the logarithm of the marginal likelihood with respect to an approximate density and hyperparameters:

$$\log p(y|x,\varphi) = \log \int p(y,\theta|x,\varphi)d\theta = \log \int \frac{p(y|\theta,x,\varphi)p(\theta|\varphi)}{q(\theta)} q(\theta)d\theta \geq$$

$$\int (\log p(y,\theta|x,\varphi) - \log q(\theta))q(\theta)d\theta =$$

$$\log p(y|x,\varphi) - \int (\log q(\theta) - \log p(\theta|y,x,\varphi))q(\theta)d\theta =$$

$$\log p(y|x,\varphi) - KL\big(q(\theta)||p(\theta|y,x,\varphi)\big) = ELBO(q,\varphi)$$

Different approximate densities might be useful in different situations

Complexity

Mean-field approximations (Independent Gaussian approximation)

- All components are independent distributions (Wainwright and Jordan (2008)).

Gaussian approximation

- Gaussian distribution (Tan, Bhaskaran and Nott (2019)).

Neural network (Normalizing flows)

- Simple distribution is passed through a neural network (Rezende and Mohamed (2015)).

Non-parametric approximation (Stein variational inference)

- Approximate functional space optimization (Liu and Wang (2019)).

Quality

Sparse Bayesian neural network for inflation forecasting

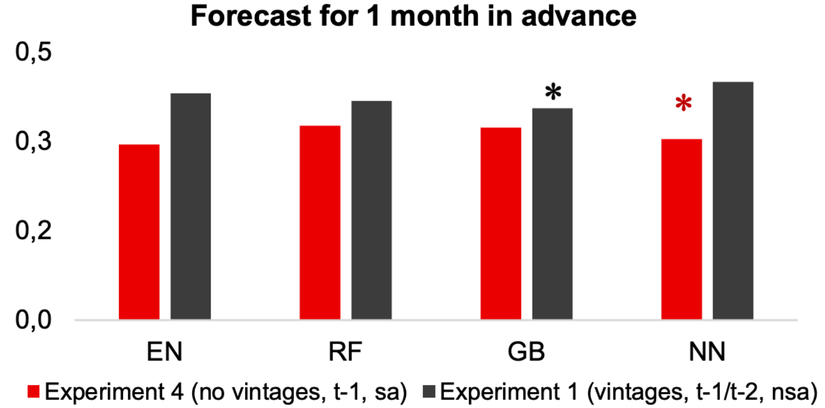
Neural network

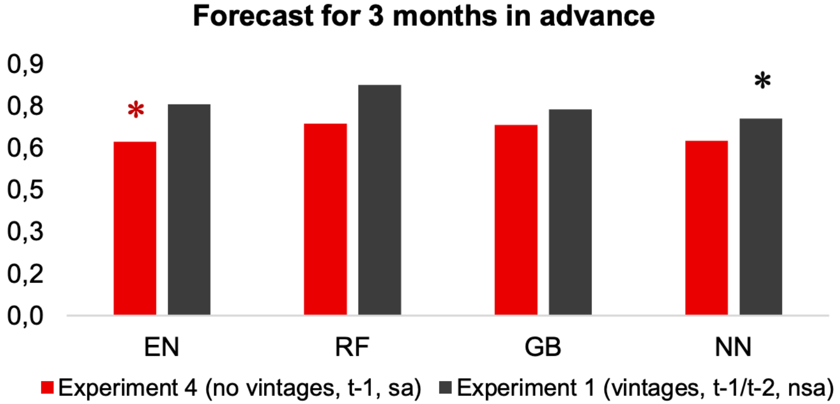
- Highly non-linear and flexible model;
- Sparse Bayesian regularization (Tipping (2001)) to avoid overfitting and minimize cross-validation dimensionality.

Model performance

- Bayesian neural network is usually comparable with or better than other ML models for inflation prediction;
- Due to a sparse structure, it can be easy to find the most important features (Khabibullin and Seleznev (2020)).

**Mean-field variational approximation is used.*





RMSFEs of inflation forecasting for different ML models developed by Mamedli and Shibitov (2021)

Seasonal adjustment model for financial flows

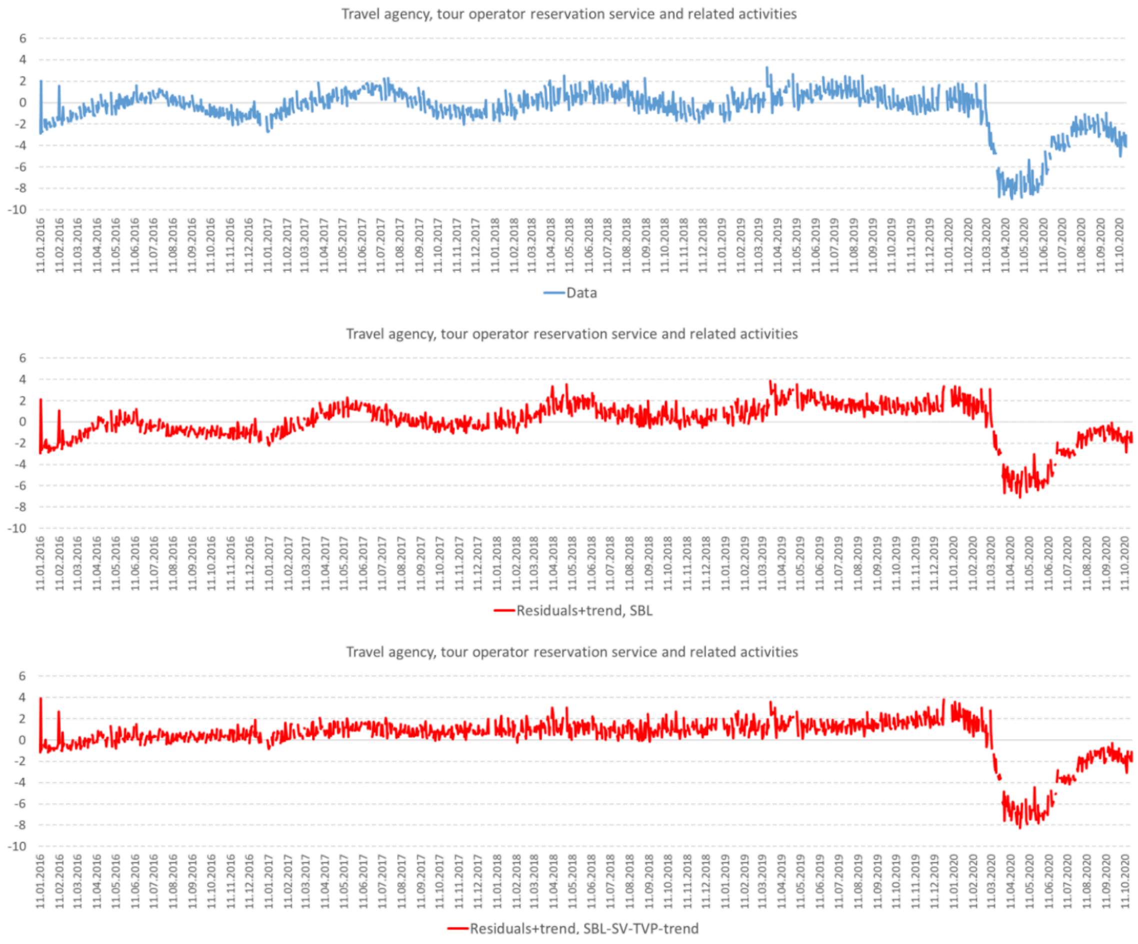
Financial flow data

- A financial flow report which helps monitor real-time industry inflows and outflows was created at the start of the COVID-19 crisis;
- Existence of daily, weakly and monthly seasonality greatly complicates the understanding of reasons behind changes in indicator dynamics.

Seasonal adjustment

- Hundreds of series are smoothed every day;
- Models without flexible trends (RW trend with SV) show poor performance for some series.

**Mean-field and normalizing flows variational approximation is used. For uncertainty estimation, Stein VI produces significantly more accurate results.*



Daily seasonal adjustment during the COVID-19 crisis using models with and without a flexible trend proposed by Khabibullin et al. (2020)

Incorporating economic judgment into nonlinear models

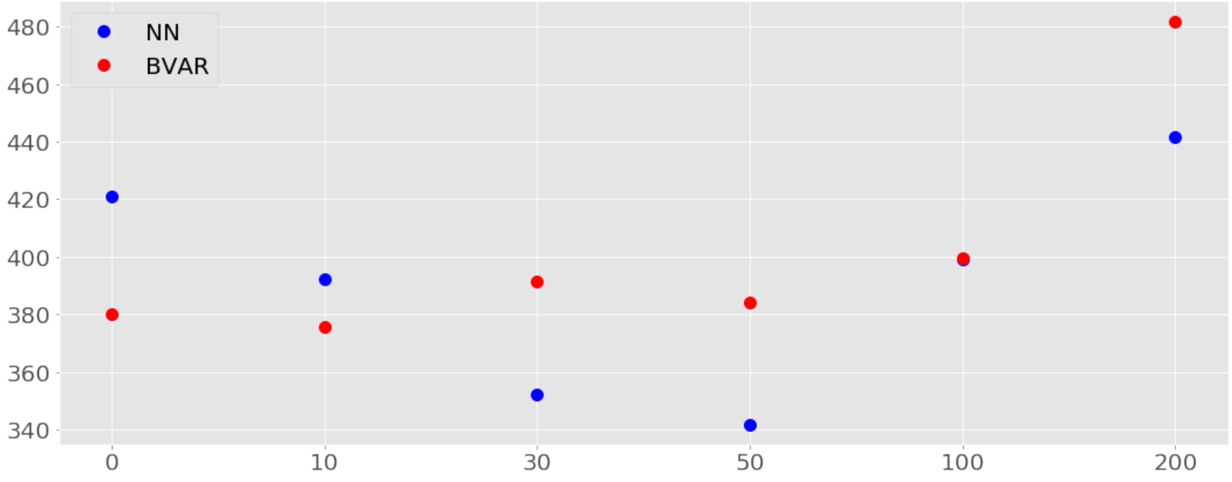
Guided non-structural model

- An extension of the idea suggested by Del Negro and Schorfheide (2004);
- Real data and artificial data from a structural model are used for the estimation of a non-structural model;
- The final algorithm for the joint structural and non-structural (hyper)parameters estimation is a double variational inference or ADVIL (Li et al. (2019)).

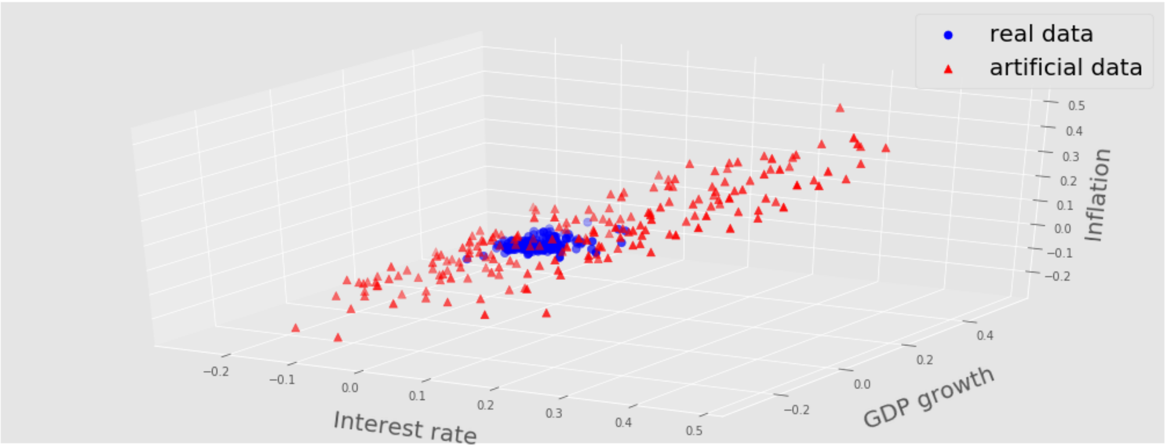
Results

- Economic judgment from the DSGE model helps to improve BNN model results.

**Mean-field and normalizing flows variational approximations are used.*



NLL for a different choice of the number of artificial points



Real and artificial points for BNN

Conclusion

- Variational inference helps in the estimation of models that cannot be estimated via classical Bayesian techniques;
- Our experience shows that even poor mean-field approximation is usually sufficient for practical applications.