



Machine learning the macroeconomic effects of financial shocks[☆]

Niko Hauzenberger^a, Florian Huber^{b,e} , Karin Klieber^{c,*}, Massimiliano Marcellino^{d,e,f,g,h}

^a University of Strathclyde, United Kingdom

^b University of Salzburg, Austria

^c Oesterreichische Nationalbank, Austria

^d Bocconi University, Italy

^e Baffi, Italy

^f BIDS, Italy

^g IGIER, Italy

^h CEPR, United Kingdom

ARTICLE INFO

JEL classification:

C11
C30
C45
E3
E44

Keywords:

Bayesian neural networks
Nonlinear local projections
Financial shocks
Asymmetric shock transmission

ABSTRACT

We propose a method to learn the nonlinear impulse responses to structural shocks using neural networks, and apply it to uncover the effects of US financial shocks. The results reveal substantial asymmetries with respect to the sign of the shock. Adverse financial shocks have powerful effects on the US economy, while benign shocks trigger much smaller reactions. Instead, with respect to the size of the shocks, we find no discernible asymmetries.

1. Introduction

Extreme financial shocks such as the bankruptcy of Lehman Brothers in September 2008 have the potential to trigger substantial nonlinear reactions on the real side of the economy. These nonlinearities appear in terms of size, sign and state dependence (Brunnermeier and Sannikov, 2014). These asymmetric effects, however, often arise from assuming particular functional relations for the conditional mean part of the model and are thus inherently model-dependent. In this note, we solve this issue by “machine learning” the nonlinear effects of financial shocks on US macroeconomic aggregates using Bayesian neural networks.

Existing studies learn the domestic effects of financial shocks either through flexible nonlinear parametric models (Barnichon et al., 2022) or through nonparametric techniques (Mumtaz and Piffer, 2022). In this note, we propose using Bayesian Neural Networks (BNNs),

a model that has, so far, demonstrated its effectiveness in forecasting (Hauzenberger et al., 2024). By basing local projections on BNNs, we gain additional flexibility in modeling nonlinearities and address the common criticism of black-box models as being unsuitable for structural economic analysis, demonstrating that BNNs can offer meaningful insights into economic dynamics.

We use BNNs as developed in Hauzenberger et al. (2024) to investigate how key US macroeconomic quantities react to a financial shock, and assess whether the reactions are non-proportional with respect to the size and asymmetric with respect to the sign of the shock. Specifically, we develop BNN-based nonlinear local projections (NLPs, see Jordà, 2005) and investigate how financial shocks – measured with the excess bond premium (EBP, Gilchrist and Zakrajšek, 2012) – impact US inflation, industrial production, and employment. We find substantial asymmetries with respect to sign, with negative shocks exerting

[☆] The authors thank Serena Ng for helpful comments and suggestions. Hauzenberger and Huber gratefully acknowledge financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (OeNB, grant no. 18763) and the Austrian Science Fund (FWF, grant no. ZK 35). Marcellino thanks for funding the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP B43C22000760006). The views expressed in this paper do not necessarily reflect those of the Oesterreichische Nationalbank or the Eurosystem or the European Union, and they cannot be held responsible for them.

* Correspondence to: Monetary Policy Section, Oesterreichische Nationalbank, Otto-Wagner-Platz 3, 1090 Vienna, Austria.

E-mail address: karin.klieber@oenb.at (K. Klieber).

stronger effects than positive ones, but substantial proportionality with respect to the size of the shock.

2. Bayesian Neural Networks (BNNs)

We follow [Hauzenberger et al. \(2024\)](#) and use Bayesian Neural Networks (BNNs) to model a macroeconomic response y_t as an unknown function of K covariates \mathbf{x}_t for $t = 1, \dots, T$. The model is given by:

$$y_t = \mathbf{x}_t' \boldsymbol{\gamma} + f(\mathbf{x}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2), \quad (1)$$

$$f(\mathbf{x}_t) \approx \hat{f}_L(\mathbf{x}_t) = \mathbf{W}_{L+1} \mathbf{h}_L(\mathbf{W}_L \mathbf{h}_{L-1}(\dots \mathbf{W}_2 \mathbf{h}_1(\mathbf{W}_1 \mathbf{x}_t))), \quad (2)$$

where $\boldsymbol{\gamma}$ denotes linear coefficients of dimension K , and ε_t is a Gaussian shock with zero mean and time-varying variance σ_t^2 .¹ We allow for L hidden layers and Q_ℓ ($\ell = 1, \dots, L$) neurons in each layer. By recursively applying nonlinear transformations to the neurons of the previous layer, we move from covariates \mathbf{x}_t to $f(\mathbf{x}_t)$, the output of the final layer. These transformations are implemented via activation function $h_{\ell,q}$ ($q = 1, \dots, Q_\ell$), which in the case of [Hauzenberger et al. \(2024\)](#) can be layer and neuron-specific (i.e., $\mathbf{h}_\ell = (h_{\ell,1}, \dots, h_{\ell,Q_\ell})'$).² The network coefficients are stored in \mathbf{W}_ℓ for $\ell = 2, \dots, L$ with dimension $Q_\ell \times Q_{\ell-1}$, in \mathbf{W}_{L+1} as a $1 \times Q_L$ vector and \mathbf{W}_1 as a $Q_1 \times K$ matrix.

To address the sharp increase in the number of coefficients with more complex network structures, we introduce regularization via a global local shrinkage prior in the form of the horseshoe prior ([Carvalho et al., 2009](#); [Ghosh et al., 2019](#); [Bhadra et al., 2020](#)). In this setup, each element of $\mathbf{w}_{\ell,i} = (w_{\ell,i1}, \dots, w_{\ell,iQ_{\ell-1}})'$, with $\mathbf{w}_{\ell,i}$ denoting the i th row of \mathbf{W}_ℓ ($\ell = 1, \dots, L+1$) follows:

$$w_{\ell,ij} \sim \mathcal{N}(0, \phi_{\ell,ij}), \quad \phi_{\ell,ij} = \lambda_{\ell,i}^2 \varphi_{\ell,ij}^2, \quad \lambda_{\ell,i} \sim C^+(0, 1), \quad \varphi_{\ell,ij} \sim C^+(0, 1).$$

The global (neuron-specific) shrinkage parameter $\lambda_{\ell,i}$ forces all elements in $\mathbf{w}_{\ell,i}$ towards zero, while the local scaling parameter $\varphi_{\ell,ij}$ allows for coefficient-specific shrinkage. Accordingly, we apply the horseshoe prior to the linear coefficients $\boldsymbol{\gamma}$.

One important specificity of the network structure is that it allows for a mixture specification, averaging over four different activation functions. We define $h^{(m)}$ as one out of $m \in \{\text{leakyReLU, sigmoid, ReLU, tanh}\}$ activation functions and let each be given by:

$$h_{\ell,q}(z_{\ell,q,t}) = \sum_{m=1}^4 \omega_{\ell,q}^{(m)} h^{(m)}(z_{\ell,q,t}), \quad (3)$$

with $z_{\ell,q,t}$ denoting the q th element in the recursively defined vector $\mathbf{z}_{\ell,t} = \mathbf{W}_\ell \mathbf{h}_{\ell-1}(\mathbf{z}_{\ell-1,t})$ and $\mathbf{z}_{1,t} = \mathbf{W}_1 \mathbf{x}_t$. Weights $\omega_{\ell,q}^{(m)}$ are constrained to satisfy $\omega_{\ell,q}^{(m)} \geq 0$ and $\sum_m \omega_{\ell,q}^{(m)} = 1$. The prior on $\omega_{\ell,q}^{(m)}$ is set in an uninformative manner with a prior probability of $\text{Prob}(\delta_q = m) = 1/4$.

For posterior inference, we use an MCMC algorithm structured into multiple blocks, iterated 20,000 times with the initial 10,000 draws discarded as burn-in. In brief, we start by drawing the linear coefficients $\boldsymbol{\gamma}$ and the nonlinear coefficients of the last layer \mathbf{W}_{L+1} jointly from a standard multivariate Gaussian posterior. The remaining coefficients \mathbf{W}_ℓ , for $\ell = 1, \dots, L$ are obtained via an HMC step ([Neal, 2011](#)). All shrinkage hyperparameters are updated by sampling from inverse Gamma distributions using the sampler as in [Makalic and Schmidt \(2015\)](#). Finally, to simulate the activation function $h_{\ell,q}$, we draw the indicator $\delta_{\ell,q}$ from a multinomial distribution. $\delta_{\ell,q}$ takes integer values from one to four, each corresponding to a specific activation function selected from the predefined set. For technical details we refer to [Hauzenberger et al. \(2024\)](#).

¹ The error variance is modeled via stochastic volatility, as in [Kastner and Frühwirth-Schnatter \(2014\)](#).

² Note that this discussion abstracts from the bias term to simplify notation. In our empirical application, we include the bias term, which in a neural network, allows the activation function to be shifted towards positive and negative values.

3. Nonlinear local projections in BNNs

In the literature, there is substantial evidence of asymmetric effects of benign versus adverse financial shocks on the economy (see, e.g., [Balke, 2000](#); [Brunnermeier and Sannikov, 2014](#); [Barnichon et al., 2022](#)). To shed light on this issue, we develop nonlinear local projections (NLPs; see, e.g., [Mumtaz and Piffer, 2022](#); [Gonçalves et al., 2024](#); [Inoue et al., 2024](#)) for BNNs.

Let ζ_t denote an exogenous instrument for a shock of interest. Adding ζ_t to our general nonlinear regression problem and iterating y_t h -periods forward (for $h = 0, \dots, H$), yields:

$$y_{t+h} = \psi_h \zeta_t + \mathbf{x}_t' \boldsymbol{\gamma}_h + \varepsilon_{t+h}' \tilde{\boldsymbol{\gamma}}_h + f_h(\zeta_t, \mathbf{x}_t, \varepsilon_{t+h}) + \varepsilon_{t+h}. \quad (4)$$

Here, for $h \geq 1$ we let $\varepsilon_{t+h} = (\varepsilon_t, \dots, \varepsilon_{t+h-1})'$ denote a h -dimensional vector of shocks for periods $t, \dots, t+h-1$, and $\tilde{\boldsymbol{\gamma}}_h$ a h -dimensional vector of associated coefficients. For $h = 0$, no shocks are included. Note that $\psi_h, \boldsymbol{\gamma}_h, \tilde{\boldsymbol{\gamma}}_h$ and f_h are horizon-specific.

We obtain the nonlinear local projections in two steps (see, e.g., [Kilian and Lütkepohl, 2017](#)). First, we obtain the NLP conditional on the full history of the data $\boldsymbol{\Omega}_t$ up to time t (which includes the instrument, past shocks and covariates) by computing the difference between the expectation of Eq. (4) given $\zeta_t = \tau$ and the expectation given $\zeta_t = 0$:

$$\text{NLP}(h, \tau, \boldsymbol{\Omega}_t) = \mathbb{E}(y_{t+h} | \zeta_t = \tau, \boldsymbol{\Omega}_t) - \mathbb{E}(y_{t+h} | \zeta_t = 0, \boldsymbol{\Omega}_t).$$

Second, this NLP depends on the observed data $\boldsymbol{\Omega}_t$. A more general way of representing the nonlinear response of the economy to a financial shock can be obtained by considering the unconditional NLP obtained as:

$$\text{NLP}(h, \tau) = \int \text{NLP}(h, \tau, \boldsymbol{\Omega}_t^r) d\boldsymbol{\Omega}_t^r, \quad (5)$$

where, as in [Kilian and Lütkepohl \(2017\)](#), $\boldsymbol{\Omega}_t^r$ is a randomly selected path of observations.

This quantity can be computed for each MCMC draw of $\psi_h, \boldsymbol{\gamma}_h, \tilde{\boldsymbol{\gamma}}_h$ and \hat{f}_h (the trained BNN approximation to f_h), yielding a posterior distribution over impulse responses to financial shocks. Specifically, for each MCMC draw we compute $\text{NLP}(h, \tau, \boldsymbol{\Omega}_t^r)$ for R different realizations of $\boldsymbol{\Omega}_t^r$ and take the mean:

$$\text{NLP}(h, \tau) \approx \frac{1}{R} \sum_{r=1}^R \text{NLP}(h, \tau, \boldsymbol{\Omega}_t^r). \quad (6)$$

Setting $R = 400$, a large value, yields a precise approximation to the integral in Eq. (5).

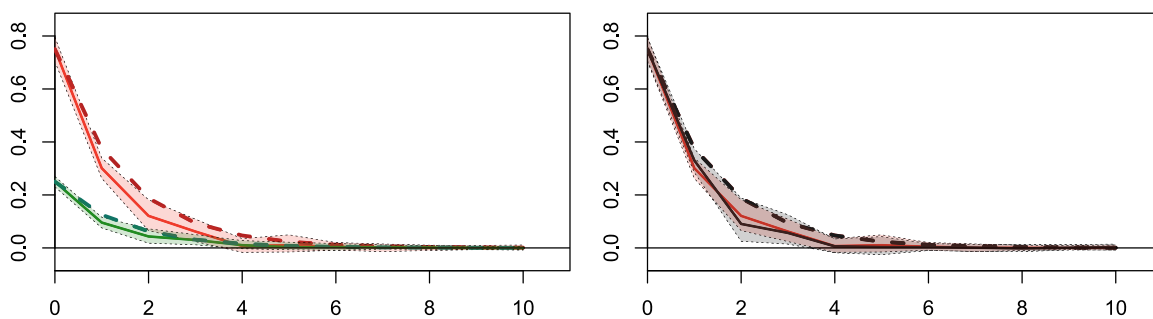
An additional complication to compute $\text{NLP}(h, \tau)$ is that ε_{t+h} in Eq. (4) is latent.³ Within a linear framework, [Lusompa \(2023\)](#) proposes estimating ε_{t+h} using the estimated residuals and then treating them as fixed regressors.

Yet, in a nonlinear framework, ignoring uncertainty surrounding ε_{t+h} can generate bias. Hence, we follow an alternative, sequential approach which exploits the Bayesian nature of our method. More specifically, in a first step we estimate the regression in Eq. (4) for $h = 0$, save the posterior distribution of the shock ε_t . Second, we estimate the model for $h = 1$, replacing ε_{t+1} in each draw of the Gibbs sampler by a draw $\varepsilon_{t+1}^{(j)} = \varepsilon_t^{(j)}$ from $p(\varepsilon_{t+1} | \bullet)$ for each t . This yields a shock distribution of $p(\varepsilon_{t+1} | \bullet)$. Third, we estimate the model for $h = 2$ replacing ε_{t+2} by $\varepsilon_{t+2}^{(j)} = (\varepsilon_t^{(j)}, \varepsilon_{t+1}^{(j)})'$ from $p(\varepsilon_{t+2}^{(j)} | \bullet)$ for each t . This procedure is repeated until we end up estimating the regression for horizon H .

³ In an extensive robustness check, [Clark et al. \(2024\)](#) show that ε_{t+h} only has a small impact on direct forecasts.

(a) Sign asymmetries (b) Size asymmetries

DGP with sign asymmetries only



DGP with size asymmetries only

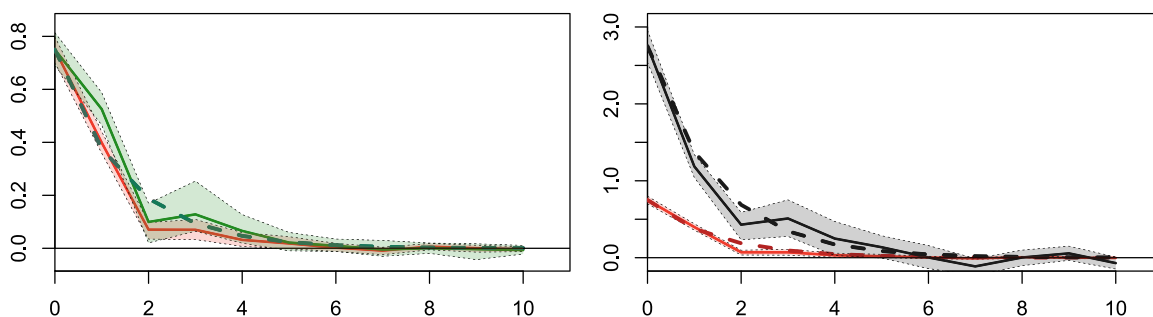


Fig. 1. Nonlinear local projections to simulated data.

Note: This figure shows asymmetries in local projections based on synthetic data simulated according to Eq. (7). The left-hand panel refers to asymmetries in the sign of the shock and the right-hand panel refers to asymmetries in the size of the shock. The solid lines indicate the posterior median, while the shaded areas refer to the 68% posterior credible interval. A positive one-unit shock is denoted in red, a negative one-unit shock is shown in green, and a positive three-unit shock is shown in gray. The dashed lines represent the associated true impulse responses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Simulation exercise

In this section we evaluate whether our approach is capable of recovering the true impulse responses to a particular shock using a highly nonlinear data generating process (DGP). Inspired by the recent nonlinear local projections literature, our stylized DGP takes the following form (see Goncalves et al., 2024):

$$y_t = \phi y_{t-1} + \beta_1 \zeta_t + \beta_2 |\zeta_t| + \beta_3 \zeta_t^3 + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 0.2), \quad y_0 = 0, \quad \zeta_t \sim \mathcal{N}(0, 0.5). \quad (7)$$

Consistent with Goncalves et al. (2024), we set the autoregressive parameter to $\phi = 0.5$. While the linear effect is $\beta_1 = 0.5$, we discriminate between two main nonlinear effects, one capturing solely size asymmetries and the other capturing only sign asymmetries, by varying β_2 and β_3 . For the former DGP, we set $\beta_2 = 0$ and $\beta_3 = 0.25$; in the latter, we set $\beta_2 = 0.25$ and $\beta_3 = 0$. Within this controlled environment, we therefore deliberately exclude ζ_t^2 , as such a squared effect would capture a mix of both sign and size asymmetries. We set the number of observations to $T = 1,000$. For each DGP, we simulate 20 realizations and estimate our model on each. We then compute the posterior summary statistics, averaged across the 20 realizations for both DGPs.

Fig. 1 shows the 16th and 84th percentiles of the posterior distribution of the LPs (shaded areas), the median (solid lines) and the true IRFs (dashed lines) for different shock signs and shock sizes. While the two panels along the main diagonal show two true impulse responses, the panels along the off-diagonal show a single true response. This captures the notion that we simulate one type of asymmetry at a time by assuming either sign asymmetries or size asymmetries, but never both. The main goal of this exercise is that, while in the diagonal panels we expect differences across the estimated impulse responses, we expect them to be identical in the off-diagonal panels. The figure

suggests that our BNN does a good job in recovering the true IRFs. While there seems to be a small bias, the shapes of the responses are very close to the true ones. This bias arises from the fact that the model we estimate is mis-specified given that we do not know the true nonlinear form of the DGP. In terms of size and sign asymmetries we find that BNNs are capable of distinguishing between both.

5. The nonlinear effects of financial shocks

We focus on how financial shocks impact inflation, output and employment, and whether these effects are symmetric and proportional. To answer these questions, we employ a BNN featuring one hidden layer and $Q = K$ neurons. Building on the existing literature (Gilchrist and Zakrajšek, 2012; Barnichon et al., 2022; Mumtaz and Piffer, 2022), we construct our dataset using variables from FRED-MD (McCracken and Ng, 2016), including inflation (CPIAUCSL), employment (CE16OV), industrial production growth (INDPRO) as well as the federal funds rate (FEDFUNDS) and stock market returns (S.P.500). In addition, we also include a proxy of the financial shock. This proxy is obtained by including the excess bond premium (EBP, Gilchrist and Zakrajšek, 2012) in a structural VAR that uses the same five variables as well as the EBP. We order the EBP measure first and obtain the structural economic shock related to this variable, which can be interpreted as a financial shock. The sample ranges from January 1960 to December 2020.

Fig. 2, again, depicts the 16th and 84th posterior percentiles of the LPs for different shock signs and shock sizes. Panel (a) shows the responses of inflation, IP and employment to a one unit contractionary financial shock (in red) and to a benign financial shock (in green). The benign shock is multiplied by -1 to simplify comparison. Panel

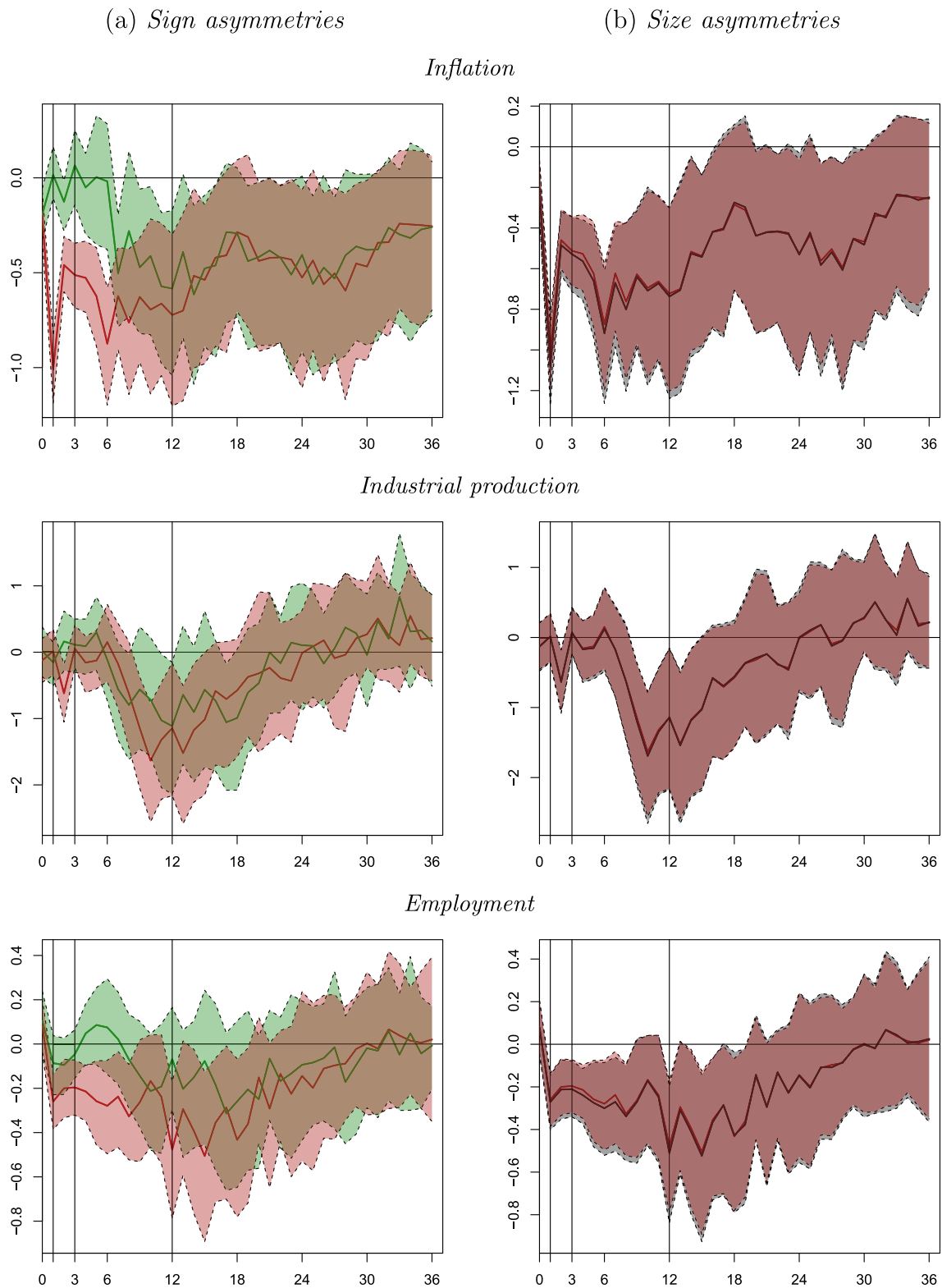


Fig. 2. Nonlinear local projections to financial shocks.
 Note: This figure shows asymmetries in local projection responses to shocks to the excess bond premium (EBP). The left-hand panel refers to asymmetries in the sign of the shock and the right-hand panel refers to asymmetries in the size of the shock. The solid lines indicate the posterior median, while the shaded areas refer to the 68% posterior credible interval. A positive one-unit shock is denoted in red, a negative one-unit shock is shown in green, and a positive three-unit shock is shown in gray. To ease comparability, we re-scale the responses associated with a negative shock and the responses associated with three-unit shock such that they represent a positive one-unit shock (multiplied by -1 and $1/3$, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(b) shows the responses to a contractionary financial shock to a one and three unit financial shock. The three unit shock is re-scaled by 1/3 to ease comparison.

The figure reveals substantial asymmetries with respect to the sign of the shock. Consistent with the literature (see, e.g., Barnichon et al., 2016, 2022; Forni et al., 2024), we find that contractionary shocks have much stronger effects on the macro aggregates than benign shocks. Starting with inflation reactions, we find that contractionary financial shocks exert downward pressure on prices. This effect is strong and peaks after around one month, with a peak median decline in inflation of around one percentage point. Notice that the reaction is also quite persistent and turns insignificant only after around 18 months. By contrast, a benign shock leads to a much weaker reaction of inflation. After around seven months (until around 1.5 years), there is some evidence that inflation picks up. But apart from this, the credible intervals mostly include zero.

We now turn to the reaction of IP growth. Output seems to react sluggishly with respect to financial shocks. The declines in output growth are much stronger if the shock is adverse, reaching almost two percentage points after around eleven months. For a benign shock, the effects are much more muted and (almost) never significant.

Similar to the inflation reaction, we find that employment growth strongly declines after a negative financial shock. This decline peters out after around two years, turning insignificant afterwards. The sharp drop in prices can be linked to the decline in employment and the associated downward pressure on wages. Again, we find no discernible reaction to a benign shock.

In terms of size asymmetries (panel (b) of Fig. 2), we find that contractionary financial shocks of different sizes (one and three units) trigger proportional reactions of all three focus variables under consideration.

6. Conclusion

We exploit Bayesian neural networks to compute non-parametric impulse response functions, and applies the method to approximate the reaction of three key US macroeconomic variables to financial shocks. We find that asymmetries arise mostly with respect to the sign of the shocks. Adverse financial shocks have a tendency to trigger much stronger reactions of inflation, industrial production and employment than benign shocks. When it comes to asymmetries with respect to size, we instead find no differences, with small and large shocks resulting in almost exactly proportional impulse responses.

Data availability

Data will be made available on request.

References

- Balke, N.S., 2000. Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Rev. Econ. Stat.* 82 (2), 344–349.
- Barnichon, R., Matthes, C., Ziegenbein, A., 2016. Theory ahead of measurement? assessing the nonlinear effects of financial market disruptions. FRB Richmond Working Paper.
- Barnichon, R., Matthes, C., Ziegenbein, A., 2022. Are the effects of financial market disruptions big or small? *Rev. Econ. Stat.* 104 (3), 557–570.
- Bhadra, A., Datta, J., Li, Y., Polson, N., 2020. Horseshoe regularisation for machine learning in complex and deep models. *Int. Stat. Rev.* 88 (2), 302–320.
- Brunnermeier, M.K., Sannikov, Y., 2014. A macroeconomic model with a financial sector. *Am. Econ. Rev.* 104 (2), 379–421.
- Carvalho, C.M., Polson, N.G., Scott, J.G., 2009. Handling sparsity via the horseshoe. In: *Artificial Intelligence and Statistics*. PMLR, pp. 73–80.
- Clark, T.E., Huber, F., Koop, G., Marcellino, M., Pfarrhofer, M., 2024. Investigating growth-at-risk using a multicountry non-parametric quantile factor model. *J. Bus. Econom. Statist.* (forthcoming).
- Forni, M., Gambetti, L., Maffei-Faccioli, N., Sala, L., 2024. Nonlinear transmission of financial shocks: Some new evidence. *J. Money, Credit. Bank.* 56 (1), 5–33.
- Ghosh, S., Yao, J., Doshi-Velez, F., 2019. Model selection in Bayesian neural networks via horseshoe priors. *J. Mach. Learn. Res.* 20 (182), 1–46.
- Gilchrist, S., Zakrajšek, E., 2012. Credit spreads and business cycle fluctuations. *Am. Econ. Rev.* 102 (4), 1692–1720.
- Goncalves, S., Herrera, A.M., Kilian, L., Pesavento, E., 2024. Nonparametric Local Projections. Federal Reserve Bank of Dallas.
- Goncalves, S., Herrera, A.M., Kilian, L., Pesavento, E., 2024. State-dependent local projections. *J. Econometrics* 105702.
- Hauzenberger, N., Huber, F., Klieber, K., Marcellino, M., 2024. Bayesian neural networks for macroeconomic analysis. *J. Econometrics* (forthcoming).
- Inoue, A., Rossi, B., Wang, Y., 2024. Local projections in unstable environments. *J. Econometrics* 105726.
- Jordà, Ò., 2005. Estimation and inference of impulse responses by local projections. *Am. Econ. Rev.* 95 (1), 161–182.
- Kastner, G., Frühwirth-Schnatter, S., 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Comput. Statist. Data Anal.* 76, 408–423.
- Kilian, L., Lütkepohl, H., 2017. *Structural Vector Autoregressive Analysis*. Cambridge University Press.
- Lusompa, A., 2023. Local projections, autocorrelation, and efficiency. *Quant. Econ.* 14 (4), 1199–1220.
- Makalic, E., Schmidt, D.F., 2015. A simple sampler for the horseshoe estimator. *IEEE Signal Process. Lett.* 23 (1), 179–182.
- McCracken, M.W., Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. *J. Bus. Econom. Statist.* 34 (4), 574–589.
- Mumtaz, H., Piffer, M., 2022. Impulse response estimation via flexible local projections. arXiv:2204.13150.
- Neal, R.M., 2011. MCMC using Hamiltonian dynamics. *Handb. Markov Chain Monte Carlo* 2 (11), 2.