Economics-Informed Neural Networks for Macroeconomic Forecasting

Hui Chen   Yuhan Cheng

MIT Sloan and NBER
Shandong University

2024 RiskLab/BoF/ESRB Conference on AI and Systemic Risk Analytics

June 7, 2024

Outline

“The End of Theory”

“All models are wrong, but some are useful.” So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? …

Today companies like Google, which have grown up in an era of massively abundant data, don’t have to settle for wrong models. Indeed, they don’t have to settle for models at all.

Chris Anderson, Editor in Chief, Wired, 2008

How to teach a robot to play pool?

How should we teach a robot to play pool?

- Data-driven approach:
  - ↪ x: shot angle, position, force ...
  - ↪ y: ball direction, speed, spin ...
  - ↪ However, labeled data are often expensive to collect.
- Theory-driven approach:
  - ↪ Model of elastic collisions
  - ↪ Conservation of linear momentum and kinetic energy ⇒ complete predictions for ball movements
  - ↪ However, the model is likely misspecified (inelastic collision, imperfect surface ...)
- How to combine the knowledge from theory and data?
  - ↪ Our answer: transfer learning
  - ↪ First learn from theory (through synthetic data), then from real data.

Challenges in Macroeconomic Forecasting

1. **Vector autoregression (VAR)**:
   - ↪ Pro: Easy to estimate, interpret;
   - ↪ Con: Does not capture non-linear dynamics; require additional structural restrictions in high dimensional cases (e.g., Bayesian VAR).
2. **DSGE models**:
   - ↪ Pro: Strong economic foundation; counterfactual analysis;
   - ↪ Con: Subject to misspecification (structural restrictions, distributional assumptions, non-stationarity); difficulty with high-dimensional features (curse of dimensionality).
3. **Lack of data**:
   - ↪ Macroeconomic data are only available at low frequency (monthly or quarterly);
   - ↪ Structural breaks, non-stationarity;
   - ↪ Sample size is too small for the training of flexible models such as DNNs.

Overview of Results

- Transfer learning significantly outperforms a DSGE-VAR model and a conventional deep learning model in macroeconomic forecasting.
- This is especially true when:
  - ↪ the size of training sample is small;
  - ↪ or the market condition is volatile (e.g., during the COVID Pandemic)
- In TL, structural restrictions help with regularization, but they are not treated as hard constraints, which would likely add more bias.
  - ↪ This differs from imposing structural restrictions strictly in ML model (e.g., PINNs).
- Loosely speaking, we can think of the structural model providing an "informative prior," but there are important differences.
- A flexible and easy-to-use framework to use theory to guide our learning from data.

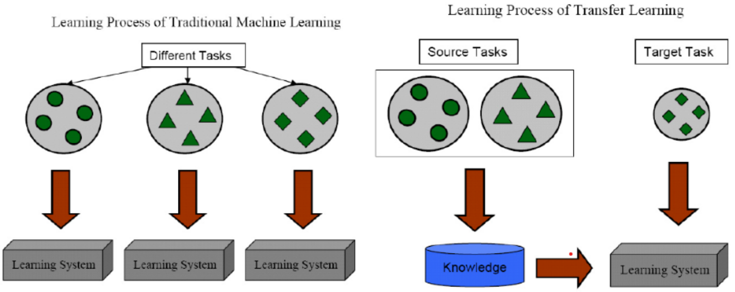
Related Literature

- Economic restrictions in ML: Garcia and Gencay (2000), Campbell and Thompson (2008), Avramov, Cheng, and Metzker (2022), Chen, Pelger, and Zhu (2023), Campello, Cong, and Zhou (2023), Chen, Cheng, Liu, and Tang (2023)
- Physics-informed neural networks: Raissi, Perdikaris, and Karniadakis (2019) ...
- Macroeconomic forecasting: Smets and Wouters (2003, 2007), Christiano, Eichenbaum, and Evans (2005), Del Negro, Giannoni, and Schorfheide (2015)
- Bayesian VAR: Doan, Litterman, and Sims (1984), Litterman (1986), DeJong, Ingram, and Whiteman (1993), Ingram and Whiteman (1994), Del Negro and Schorfheide (2004) ...
- ML applications in economics and finance: Hutchinson, Lo, and Poggio (1994), Chen and White (1999), Chen and Ludvigson (2009), Kelly and Pruitt (2013), Rapach and Zhou (2013), Chinco, Clark-Joseph, and Ye (2019), Jang and Lee (2019), Scheidegger and Bilionis (2019), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), Bryzgalova et al. (2021), Martin and Nagel (2022), Chen, Didisheim, and Scheidegger (2023), Goulet Coulombe et al. (2022)

Outline

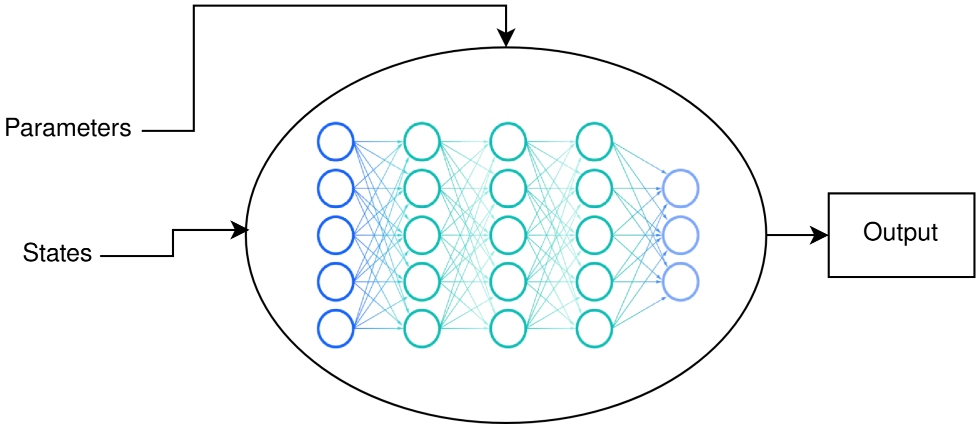
Transfer learning

- Transfer Learning: Using pre-trained models on large datasets (source domain) to accelerate learning on a smaller dataset (source domain) for a specific task.
  - ↪ computer vision: pre-train CNNs on ImageNet ⇒ medical image analysis
  - ↪ natural language processing: pre-trained LLMs ⇒ sentiment analysis
- In our setting: Source domain = economic model; target domain = actual data.



(a) Traditional Machine Learning (b) Transfer Learning

Deep surrogate: "look-up table" in the age of AI



$$\phi(\mathbf{x}_t|\theta_{NN}) = y_t$$

Learning from theory

Chen, Didisheim and Scheidegger (2023)

- What does "learning from theory" mean?
  - ↪ Structural model = a set of economic restrictions.
  - ↪ Inherit the structural restrictions.
- The advantages of deep surrogates:
  - ↪ Knowledge of true DGP: Generate as much training data as desired with (essentially) no error.
  - ↪ Expressivity: Universal approximation theorem for shallow and deep networks (Hornik, Stinchcombe, and White 1989, Hanin and Sellke 2017, Lu et al. 2017).
  - ↪ Curse of dimensionality: With suitable target function and activation function, can train accurate surrogate with sample sizes that grow polynomially (vs. exponentially) in the dimensionality of the model (Berner et al. 2020).

Source domain

- Generate synthetic data $\left((x_i, \hat{y}_i)\right)_{i=1}^{m}$ from the model-implied joint distribution, $\mathbb{P}_{(\mathscr{X}, \hat{\mathscr{Y}})}$.
  - Represent model-predicted equilibrium quantities by $\hat{y} \equiv E[y|s,h] = F(s,h|\theta)$, where s, h represent observable and hidden states. Define the augmented state $x \equiv (s, h, \theta)$.
  - Impose a hierarchical prior on θ to derive a joint probability distribution of x and $\hat{y}$, $\mathbb{P}_{(\mathscr{X}, \hat{\mathscr{Y}})}$.
- Denote a neural network with L layers as $\Phi(\sigma_1, \cdots, \sigma_L; W_1, \cdots, W_L)$.
  - σ_i: activation function; W_i: weights
- In the source domain, we train the neural networks to learn from the synthetic data generated by the theoretical model. Effectively, we try to predict $\hat{y}_i$ using x_i:

$$\hat{\Phi} = \arg\min \left\{ \frac{1}{m} \sum_{i=1}^{m} \mathscr{L}(\Phi(x_i), \hat{y}_i) : \Phi \in \mathscr{H}_L \right\}$$

- In the application, we use MAE as the loss function.

Target domain

- In the target domain, train with real data $\left((s_i, y_i)\right)_{i=1}^{n}$.
- How to deal with hidden states h and parameters θ in x:
  - ↪ Conditioning down ($x \Rightarrow s$) or filtering ($s \Rightarrow x$).
  - ↪ With the "deep surrogate" methodology from Chen, Didisheim and Scheidegger (2023), we can use the network trained in the source domain to efficiently filter h_i and θ_i.
- Training in the target domain follows a fine-tuning method with lower learning rate and smaller number of epochs.
- Using the network architecture inherited from the source domain, we minimize empirical loss

$$\widetilde{W}_1, \cdots, \widetilde{W}_L = \arg\min \frac{1}{n} \sum_{i=1}^{n} \mathscr{L}\left(\Phi(x_i), y_i\right),$$

  using the $\widehat{W}_1, \widehat{W}_2, \cdots, \widehat{W}_L$ from the source domain as the starting point.
- Alternative approaches:
  - ↪ replacing part of the network from the source domain with new layers;
  - ↪ fixing early part of the network (frozen layers)

Teaching economics to the machines

- The neural networks trained on the simulated data from a theoretical model inherit the structural restrictions from the theory.
- Valid restrictions can help regulate learning from real data.
  - ↪ Bias-variance tradeoff
- The idea: Do not strictly impose the theoretical restrictions. Let them guide/regulate the training of the ML model on real data.
- Potential benefits:
  - ↪ Variance reduction
  - ↪ Speed up training on real data; less demand for real data
  - ↪ Improve generalizability beyond training data boundaries

Two related approaches

- Impose equilibrium conditions when estimating structural parameters or training model.
  - ↪ Rational expectations econometrics (Saracoglu and Sargent 1978, Hansen and Sargent 1980)
  - ↪ Physics-informed neural networks (Raissi, Perdikaris, and Karniadakis 2019)
- Bayesian VAR: Deriving informative priors from an economic model to estimate VARs for macro variables.
  - ↪ Random walk (Doan, Litterman, and Sims 1984; Litterman 1986)
  - ↪ DSGE models (DeJong, Ingram, and Whiteman 1993, Ingram and Whiteman 1994, Del Negro and Schorfheide 2004)
- Del Negro and Schorfheide (2004): DSGE-VAR
  - ↪ simulate time-series data from a DSGE model (with a hierarchical prior on the DSGE model parameters);
  - ↪ fit a VAR to the simulated data to form a prior for the VAR parameters;
  - ↪ derive the posterior using real data.
- Our contribution: "Generalizing" the approach to nonlinear models via transfer learning.
  - ↪ It is also an average of the simulated and real data, but the weights depend on the hyper-parameters (learning rate, epochs).

Outline

FRBNY DSGE

- The New York Fed DSGE Model: A medium scale New Keynesian model with financial frictions.
- Based on Smets and Wouters (2007), Christiano, Eichenbaum, and Evans (2005), Del Negro, Giannoni, and Schorfheide (2015)
- It builds on the neo-classical growth model by adding nominal wage and price rigidities, variable capital utilization, costs of adjusting investment, habit formation in consumption, and credit frictions (a la Bernanke, Gertler, and Gilchrist 1999).

FRBNY DSGE

The model economy includes eight classes of agents:

1. Households that consume and supply differentiated labor.
2. Labor aggregators that combine labor from individual households.
3. Final good-producing firms that aggregate intermediate goods into a final product.
4. Intermediate good-producing firms in a monopolistically competitive market.
5. Capital producers that convert final goods into capital.
6. Entrepreneurs who buy capital with internal and borrowed funds and rent it to intermediate good firms.
7. A representative bank that collects deposits from households and lends to entrepreneurs.
8. A government with a monetary authority setting short-term interest rates and a fiscal authority managing public spending and taxes.

Empirical Setting

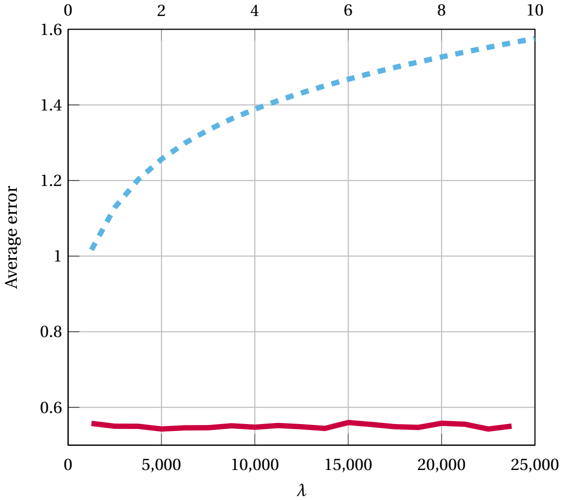
1. 13 macroeconomic variables: real output growth (including both GDP and GDI measures), consumption growth, investment growth, real wage growth, hours worked, inflation (measured by core PCE and GDP deflators), short- and long-term interest rates, 10-year inflation expectations, credit spreads, and total factor productivity.
2. Sample period: 1982 – 2023.
3. Use data from $t-5$ to t to forecast one quarter ahead, $t+1$.
4. Training with a 10-year rolling window.
5. Average of standardized errors across different macro variables:

$$WMAE = \frac{\sum_{i=1}^{p} \frac{MAE_i}{STD_i}}{p}$$

where STD_i represents the standard deviation for the i-th variable in the training sample.

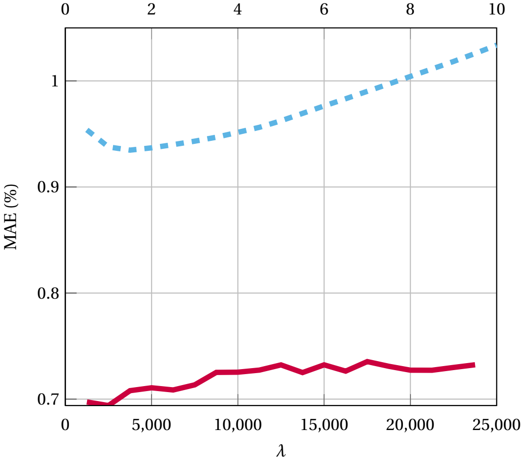
Outline

Comparison of TL and DSGE-VAR

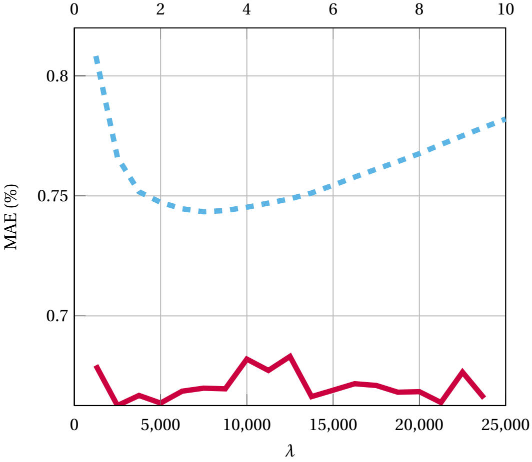


Average Performance

Comparison of TL and DSGE-VAR

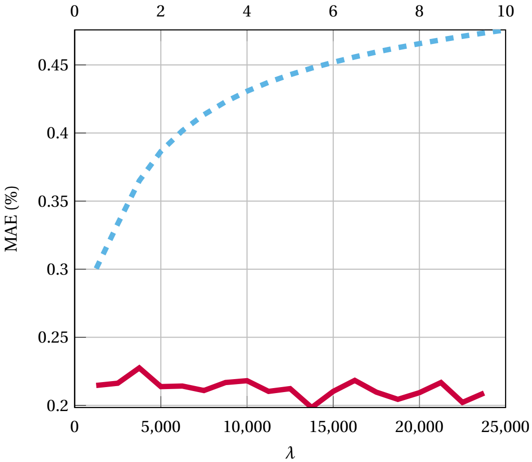


GDP growth

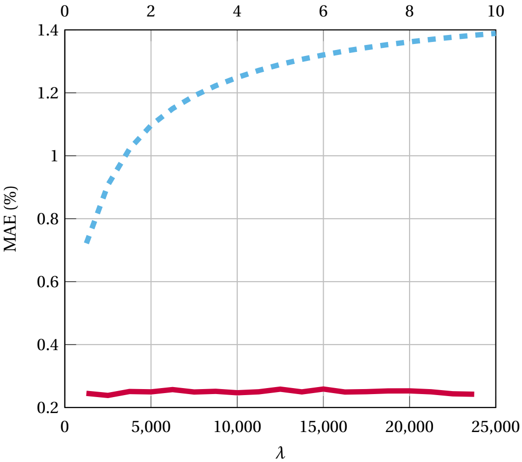


Consumption growth

Comparison of TL and DSGE-VAR

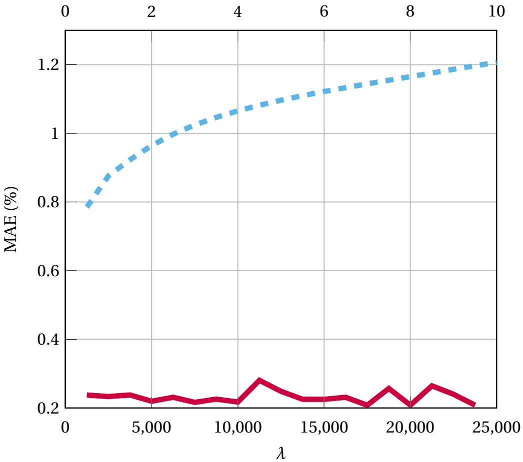


Core PCE Inflation

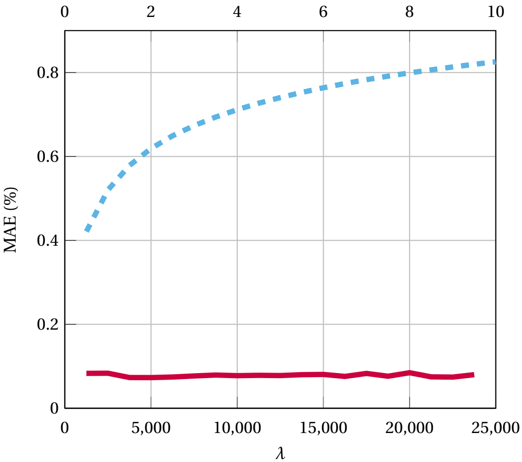


GDP Deflator Inflation

Comparison of TL and DSGE-VAR

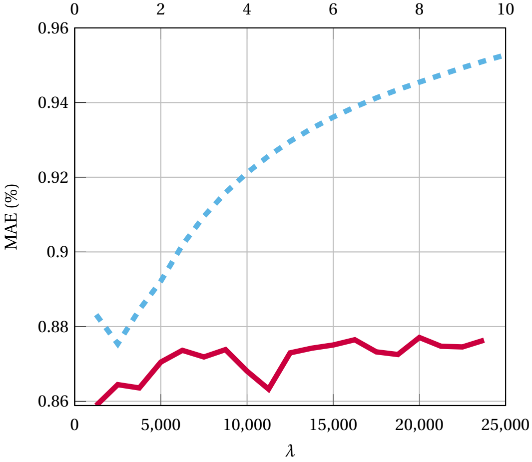


Short-term Rate

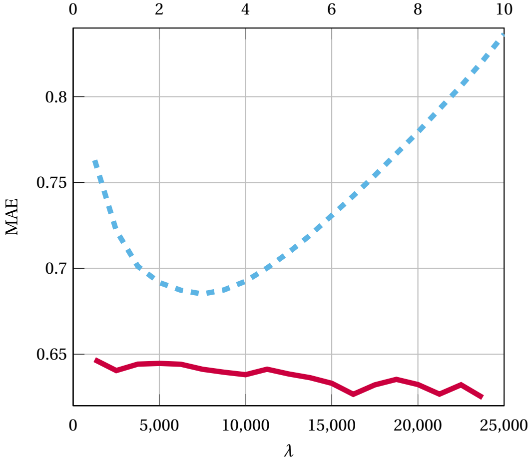


Credit Spread

Comparison of TL and DSGE-VAR

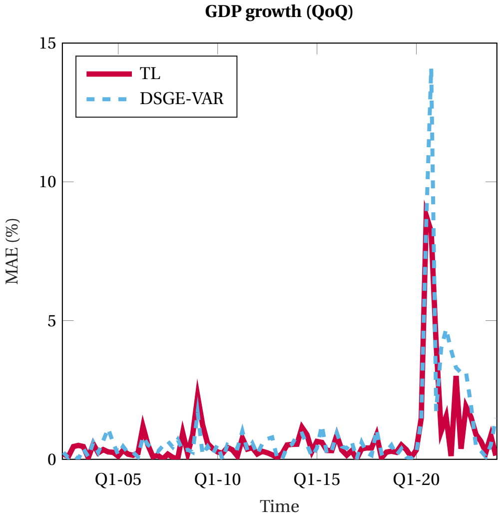


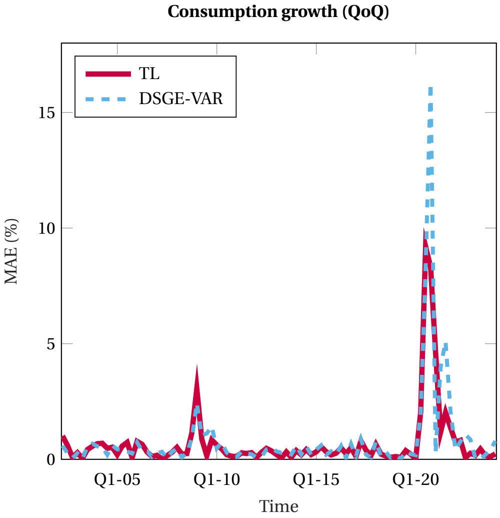
Real Wage growth



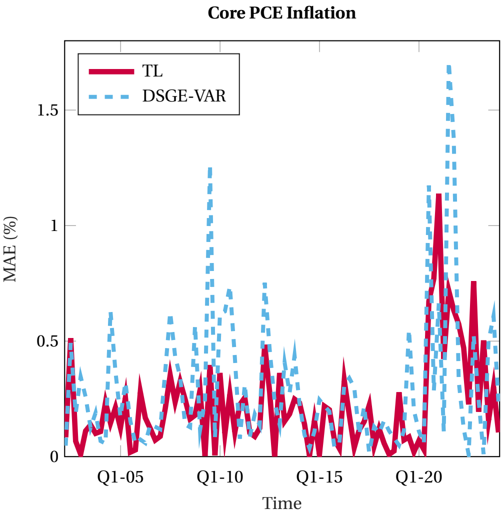
Hours

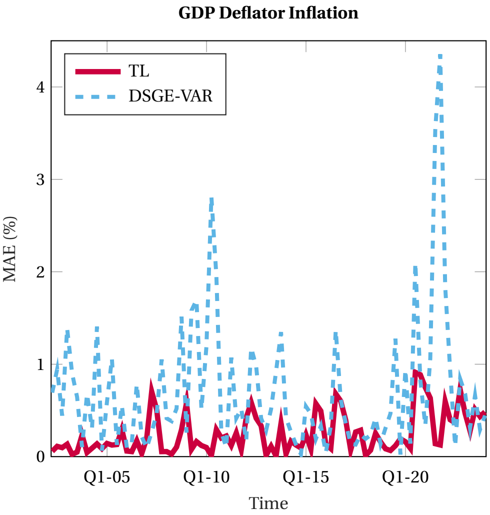
Time Series Comparison



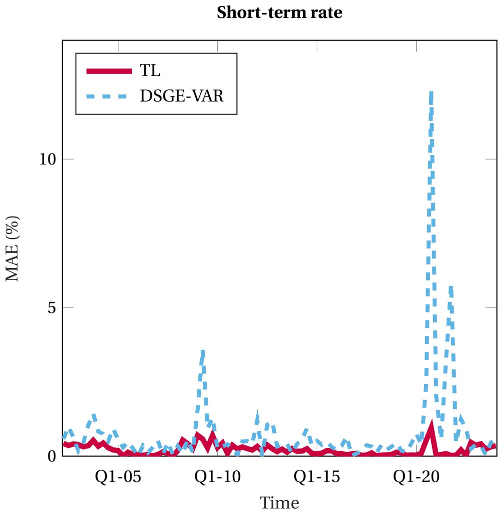


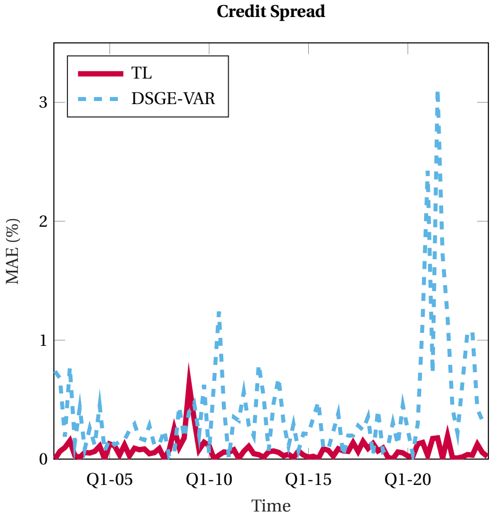
Time Series Comparison





Time Series Comparison





Why Transfer Learning Beats DSGE-VAR: Non-linearity

Variables	Coefficient	Description
distance	-1.31×10^{-8} (-0.32)	Mahalanobis distance between training set and each test point
covid	1.04^{***} (2.96)	Impact from COVID-19
crisis	0.183 (0.47)	Financial crisis
ZLB	0.607 (2.24)	Short-term interest lower than 0.1%
Δr	-0.372 (-2.55)	Rate of change of short-term interest rates
unemployment	-0.321^{***} (-3.22)	Unemployment rate
PCH	-3.08^{***} (-2.70)	Chicago Fed National Activity Index: Personal consumption and housing
PI	-0.78 (-0.87)	Chicago Fed National Activity Index: Production and income
ADS_Index	0.603^{***} (3.59)	Aruoba-Diebold-Scotti Business Conditions Index
IP	-0.103^{***} (-3.52)	Industrial production index
M2	-0.305 (-1.58)	Money supply: first difference of the log U.S. M2 money stock
EPU	-0.362 (-1.75)	Economic policy uncertainty
R^2	0.547	

Can TL Help Improve Theory?

1. How much modification is needed for DSGE models to closely approximate the true Data Generating Process in the perception of neural networks?
2. Compare the function distance before and after fine-tuning of neural networks as a proxy for the discrepancy between DSGE models and the actual data-generating process.
3. Utilizing the framework of transfer learning and the distance computation method we described, we can vary different models and broadly discuss the perceived distance between theoretical models and real data structures as understood by nonlinear artificial intelligence models.

Outline

Conclusion

1. Introduction of non-linear estimations via transfer learning to overcome limitations of traditional VAR models in macroeconomic analysis.
2. Modifications to align DSGE models with real-world, non-linear data-generating processes, enhancing predictive accuracy and relevance.
3. Transfer learning models significantly outperform DSGE-VAR models, and Demonstrate lower prediction errors and stronger resilience to economic shocks.
4. AI-enhanced models show superior adaptability and accuracy in handling complex economic phenomena. Effective management of small sample challenges in macroeconomic studies.