A Neural Phillips Curve & a Deep Output Gap

Philippe Goulet Coulombe

goulet_coulombe.philippe@uqam.ca

Université du Québec à Montréal

March 10, 2022

Context

- Inflation is back ~~(in the news)~~
- Key macro equation to think about it is the New Keynesian Phillips curve (NKPC).
- g_t is arguably the quintessential summary statistic of the macroeconomy to forecast inflation (π_t)
- g_t is also the part of π_t that the monetary authority can do most about, at least in the short/medium run.
- (Although things have changed a bit with forward guidance.)

Pain Points

- Many problems plague the estimation of the NKPC. Amongst them is the hurdle that the two key components, inflation expectations ($E_t(\pi_{t+1})$) and the output gap (g_t), are both unobserved.
- Traditional remedies include creating reasonable proxies for the notable absentees, or extracting them via some form of assumptions-heavy filtering procedure
- Then, throw $\hat{g}_t$ in a "second-stage" PC regression, and find what we usually find, i.e., pretty much nothing.

Some ML perspectives

- Essentially, the traditional paradigm is all unsupervised learning.
- That is, we extract an g_t that is *not* explicitly designed (and estimated) so to explain π_t, but rather based on some loose assumptions about its time series properties (for most of literature).
- Or we extract a common factor which can explain many of the real activity variables. But one's got to choose very wisely (without any recommandation from theory) what is included, because it typically changes g_t non-trivially.
- Can we write a *supervised* extraction problem that obliviates most of those problems?

Deep Learning to the rescue

- I move towards a supervised extraction of inflation key drivers by developing a new architecture coined *Hemisphere Neural Network*.
- Its peculiar structure allows the interpretation of the last layer's cells output as key macroeconomic latent states.
- Benefits:
  1. Nonlinearities are trivially allowed for
  2. High-Dimensionality is not a problem
  3. Can estimate the gap and its time-varying coefficient in one procedure
  4. Computations are quick and done within standard deep learning software
  5. The "black box" suddenly has an economic interpretation
  6. Good forecasts

NNKPC

- NKPC:

$$\pi_t = \theta_t E_t(\pi_{t+1}) + \gamma_t g_t + \nu_t$$

- Define expectations less stringently as $\mathcal{E}_t^{\pi}$
- Empirically, energy prices matter a lot, and may impact π_t directly

$$\pi_t = \theta_t \mathcal{E}_t^{\pi} + \gamma_t g_t + \zeta_t p_t^e + \nu_t$$

- Make this a predictive problem (what we utilmely care about)

$$\pi_{t+h} = \theta_t \mathcal{E}_t^{\pi} + \gamma_t g_t + \zeta_t p_t^e + \nu_{t+h}$$

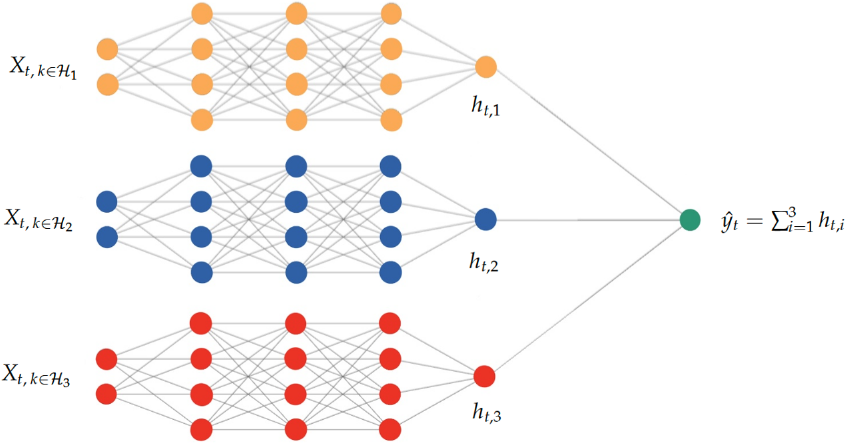
  For now, we only consider $h = 1$.
- Essentially a 3 factor model, where

$$h_{t,1} = \theta_t \mathcal{E}_t^{\pi} \qquad h_{t,2} = \gamma_t g_t \qquad h_{t,3} = \zeta_t p_t^e$$

.

Hemisphere Neural Network

- Let $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ be the expectations , real activity, and energy prices hemispheres, respectively.
- The HNN is essentially a restricted fully-connected NN.



Practical Aspects & Defining $\mathcal{H}$'s

- I split expectations in two groups: long run and exogenous ($\mathcal{E}_t^{\pi^{LR}}$), and short run ($\mathcal{E}_t^{\pi^{SR}}$) .
- Summary:

$\mathcal{H}$	Content
$\mathcal{E}_t^{\pi^{LR}}$	t (exogenous time trend)
$\mathcal{E}_t^{\pi^{SR}}$	Inflation expectations from SPF, and Michigan Survey, lags of π_t, lags of many different prices indexes in FRED
g_t	Labor Market Variables, Industrial Production Variables, National Accounts
p_t^e	Oil price, Commodities PPI, Metals PPI

Answers to FAQs

- Can Hemispheres overlap? Yes. That only changes the meaning of $h_{j,t}$'s.
- How to think of HNN and its statistical validity? There is number of proofs of DNN's nonparametric consistency for generic architectures – for instance (Farrell et al., 2021). HNN is a restricted DNN, or, one could say, a semiparametric model. If restrictions are (approximately) true, then we can be confident our $h_{j,t}$'s are close to true latent states.
- Can those restrictions be "tested"? Yes: comparing forecasting performance to a fully connected DNN.
- Need $h_{j,t}$'s be orthogonal? No.

An Important Remark

- HNN does not give us g_t, but $h_{t,g} = \gamma_t g_t$, the contribution of real activity to π_{t+1}. Ultimately, isn't it what we truly care about?
- This is *not* the neural network's doing, but the design of the problem.
- If γ_t is not time-invariant and g_t is unknown, those objects cannot be separately identified, unless we bring in some additional assumptions.
- Traditionally, g_t is treated as known in a PC regression. But clearly, γ_t's path depends directly on the postulated path of g_t.
- One possible factorization is $h_{t,g} = f_\gamma(t) f_g(\mathcal{H}_g / t)$. This means the PC coefficient is coerced to move exogeneously and slowly – like what is assumed by random walk coefficients.
- Yet, splitting the atom could be informative.

HNN-F

- Enriched Architecture for **HNN-F** (F for factorized), has a final layer

$$\begin{aligned}\hat{\pi}_{t+s} = \ & h_{\mathcal{E}^{\pi}_{\mathrm{LR}}}(t) \\ & + h_{\theta}(t) h_{\mathcal{E}^{\pi}_{\mathrm{SR}}}(\mathcal{H}_{\mathcal{E}^{\pi}}/t) \\ & + h_{\gamma}(t) h_{g}(\mathcal{H}_{g}/t) \\ & + h_{\zeta}(t) h_{p^e}(\mathcal{H}_{p^e}/t) \\ & + \nu_{t+s}.\end{aligned} \tag{1}$$

  where time-varying coefficient hemispheres outputs θ_t, γ_t and ζ_t are all forced to be non-negative (through an absolute value layer) for identification needs.

- This gives us the desired γ_t and g_t.
- Uncertainty Quantification: via "out-of-bag" block-subsampling.

Deep Dive: Tuning Parameters

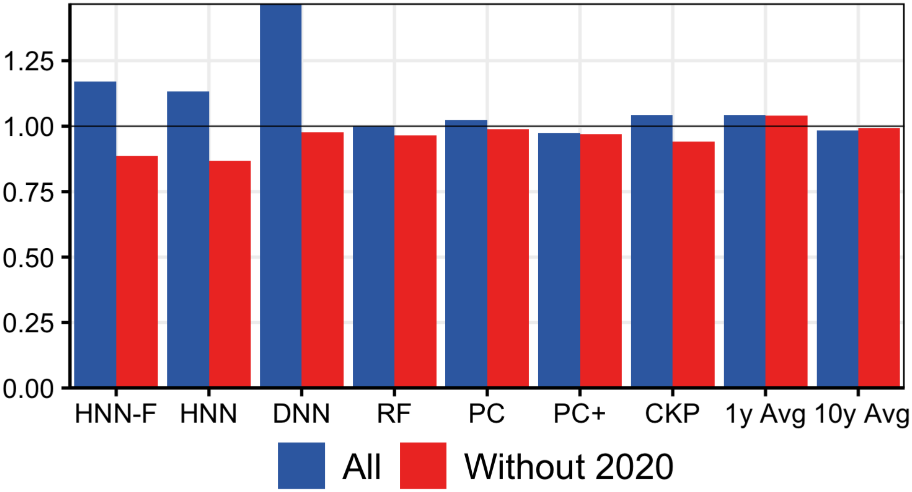
- **HNN**: each $\mathcal{H}$ is allocated 5 `layers` of 400 `neurons` each, with weight sharing.
- **HNN-F**: states hemispheres are given `neurons` $= 400$ and `layers` $= 3$ while the coefficients hemispheres (with only input being t) have `neurons` $= 100$ and `layers` $= 3$.
- Activation functions are *ReLu*, for rectified linear unit:

$$\text{ReLU}(x) = \max\{0, x\}$$

- Maximal number of `epochs` is 500. Early stopping is used based on validation MSE.
- `learning rate` is 0.00025 for HNN and 0.05 for HNN-F.
- `dropout rate` is 0.2
- `sampling rate` is 0.85, B (number of bootstraps) is 300
- `block size` is 6 quarters
- $B = 50$ is more than enough for forecasting purposes.
- $B = 300$ takes $\sim$1 hour on a M1 Macbook Air.

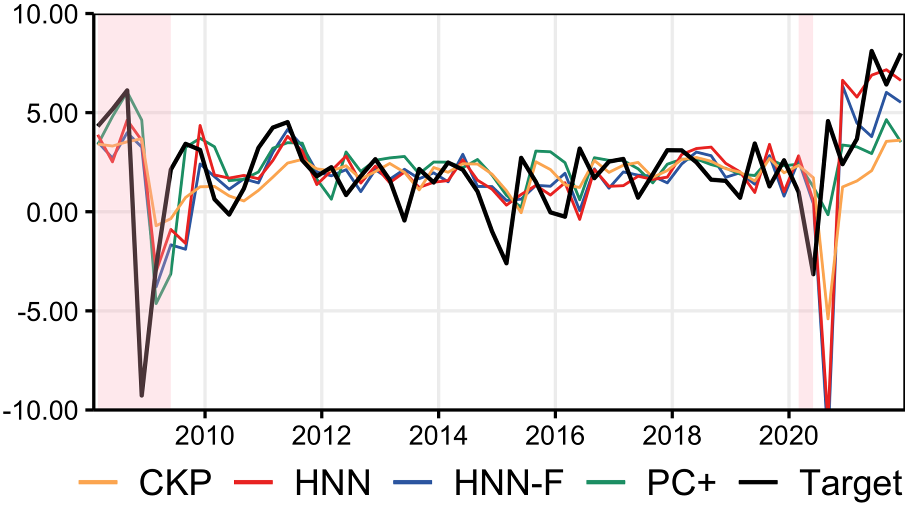
Forecasting Performance (MSEs wrt AR(4))

Benchmark Quartely data, one quarter ahead, from 2007

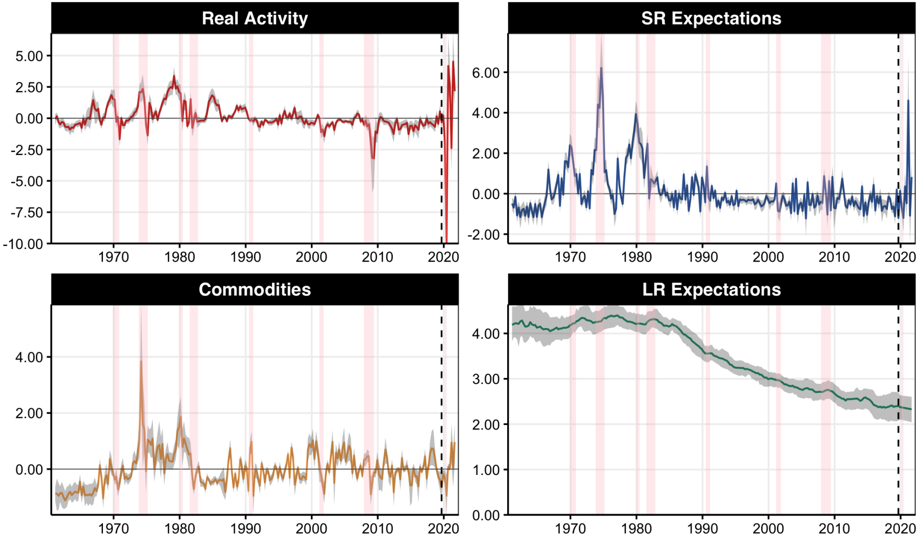


Forecasts

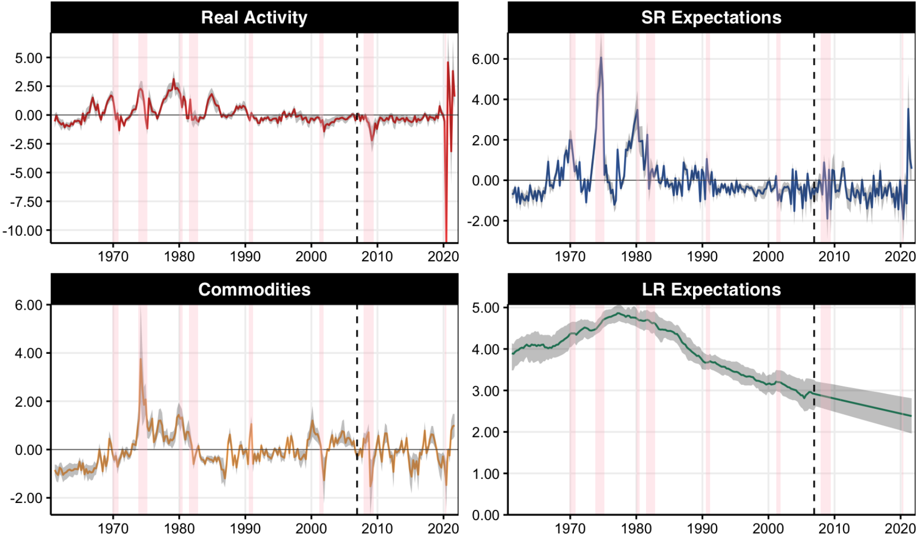
Benchmark Quartely data, one quarter ahead, from 2007



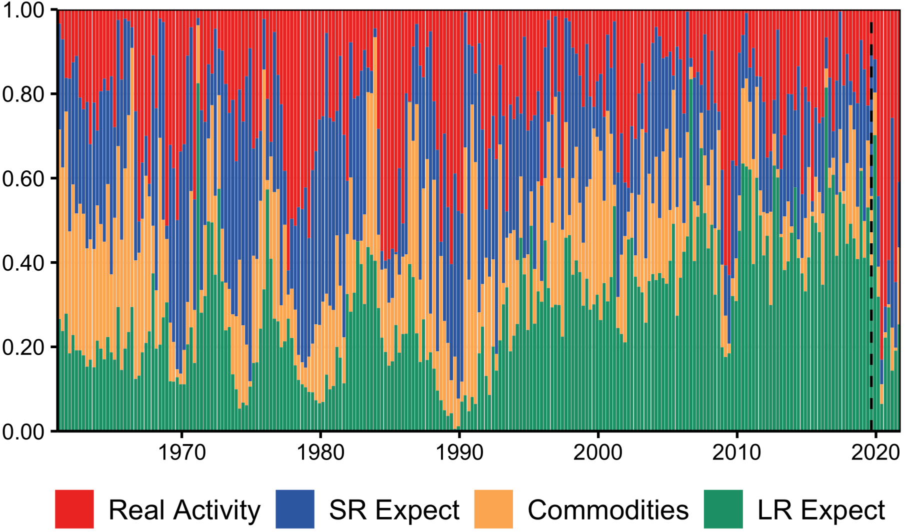
A Look at Components (h_t's)



Is History Being Rewritten?



Jackson Pollock, or inflation shares



A More Familiar View

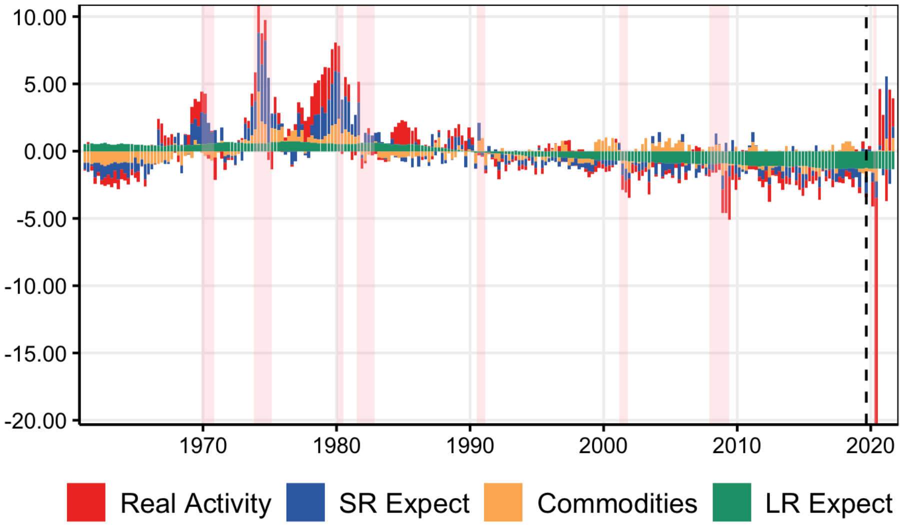


Figure: $\hat{\pi}_t$ decomposition

Comparing with Classic Approaches I

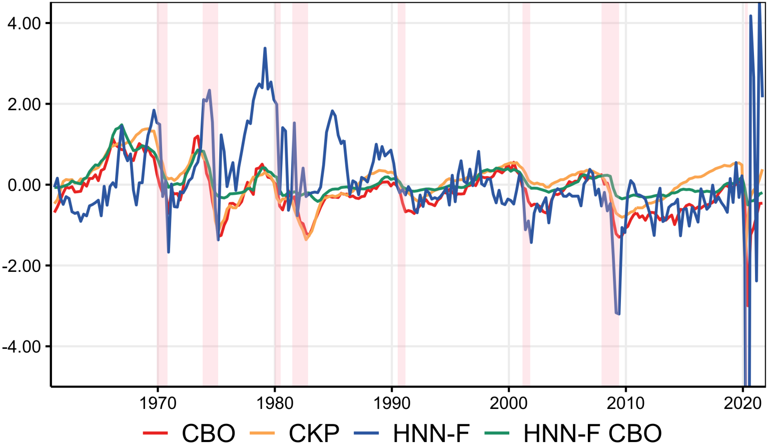


Figure: Contributions of economic activity.

- HNN-F and HNN find a substantially more important role for real activity than classic approaches.
- HNN-F is the only one showing strong overheating episodes past 2020.

Comparing with Classic Approaches II

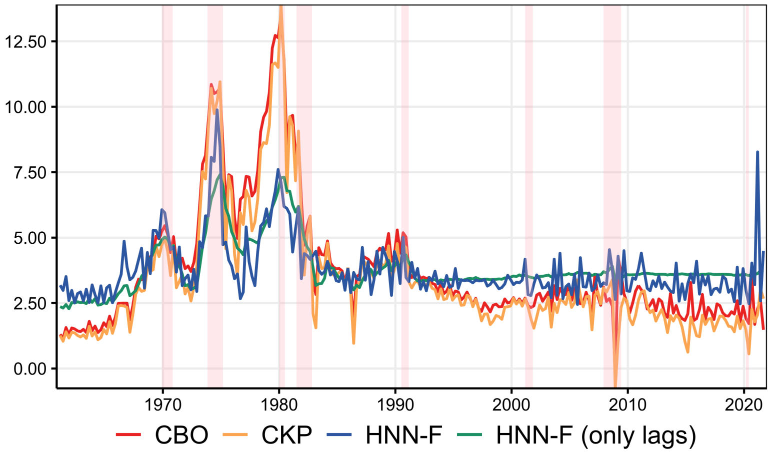


Figure: Contributions of expectations/unit costs.

- HNN-F (and HNN as well) finds a milder roles for expectations.
- HNN-F is the only one showing a flash disanchoring of short-run expectations in 2021.

Zooming on the Gap and its Coefficient

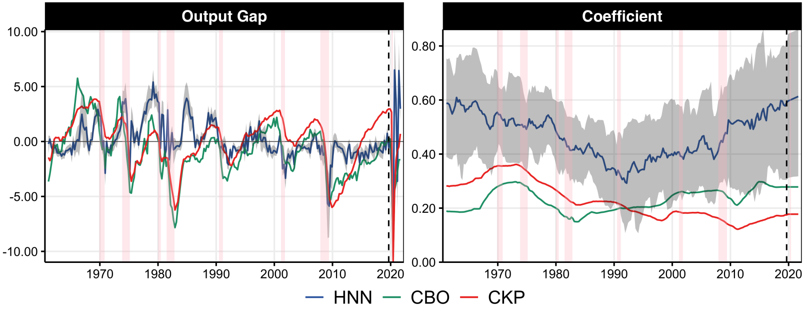
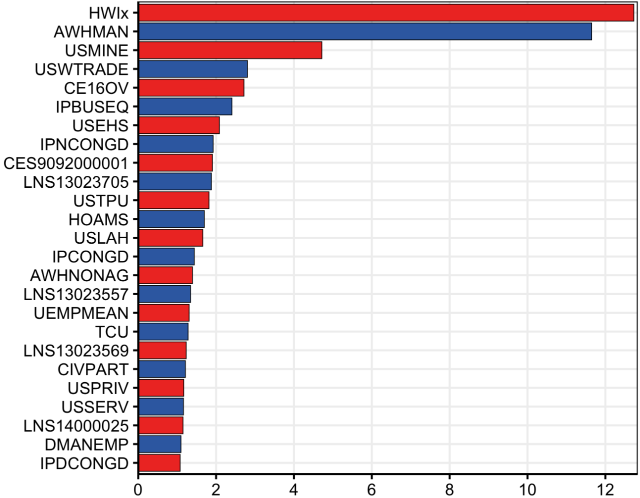


Figure: HNN-F's output gap (g_t) and associated coefficient (γ_t). Notes: Dashed line is the beginning of the out-of-sample. NBER recessions are in pink shadowing.

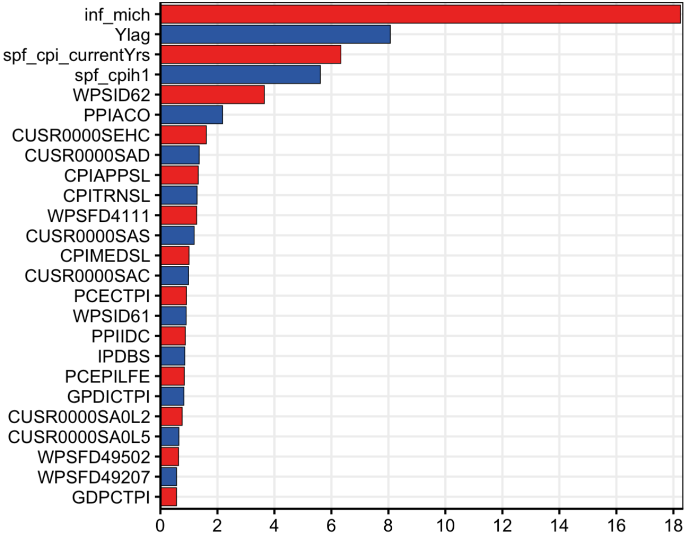
- In partial agreement with the recent literature (γ_t decreased but...).
- Unlike results from standard approaches, γ_t is not found to decline further following 2008, but rather to increase gently.
- Inflation did not go through the roof because g_t and $\mathcal{E}_t^{SR}$ spikes are "isolated".

What is g_t made of?



- AWHMAN's (average weekly hours in the manufacturing sector) predominance suggest an important for the *intensive* margin, whereas typical gap measures are mostly about extensive margin (like filtered unemployment)
- HWIx (help wanted index): the obvious things matter.
- GDP and associated measures seem unimportant, so is the unemployment rate.

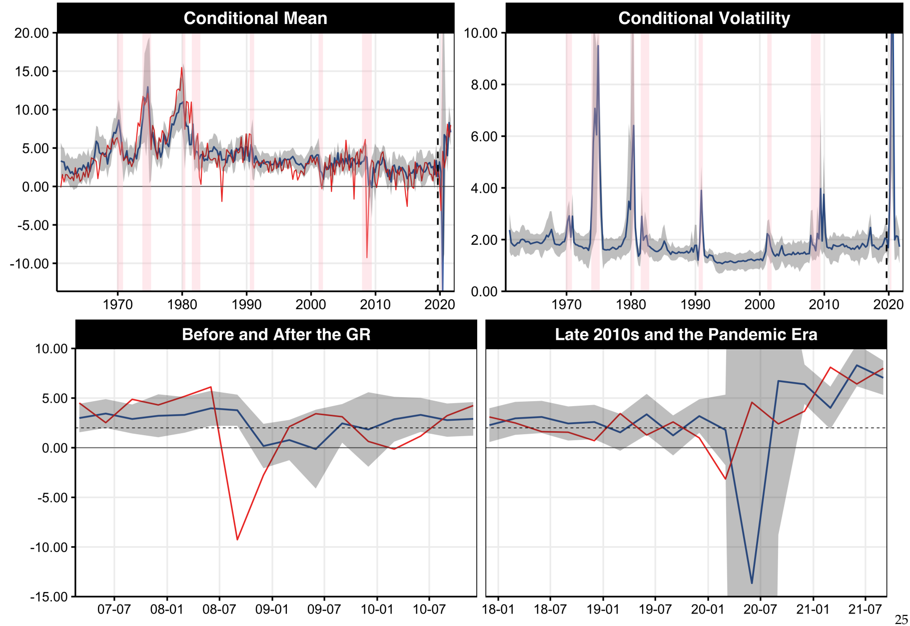
What is $\mathcal{E}_t^{\pi^{SR}}$ made of?



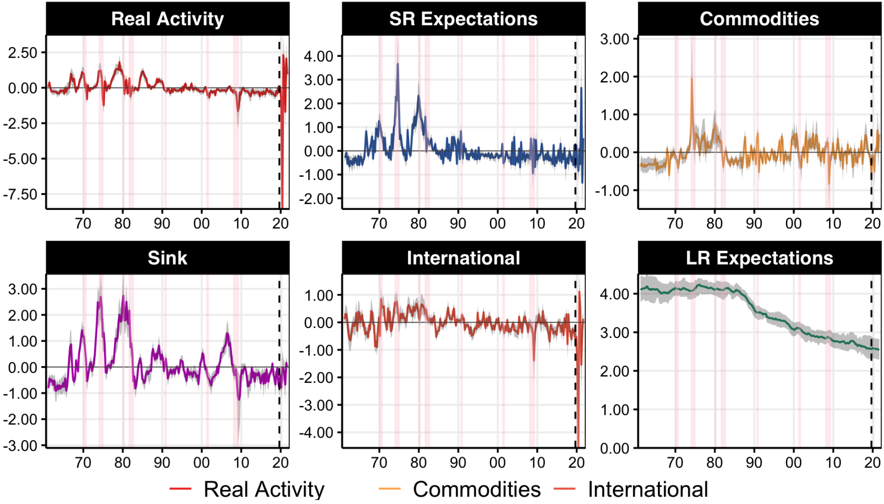
- Strengthen the case for the increasingly popular practice using of using survey expectations in PC regressions. But VI suggests including more than one seems more appropriate.
- Commodities prices, intermediate good prices, and unsurprisingly, the PPI, all matter

Encore

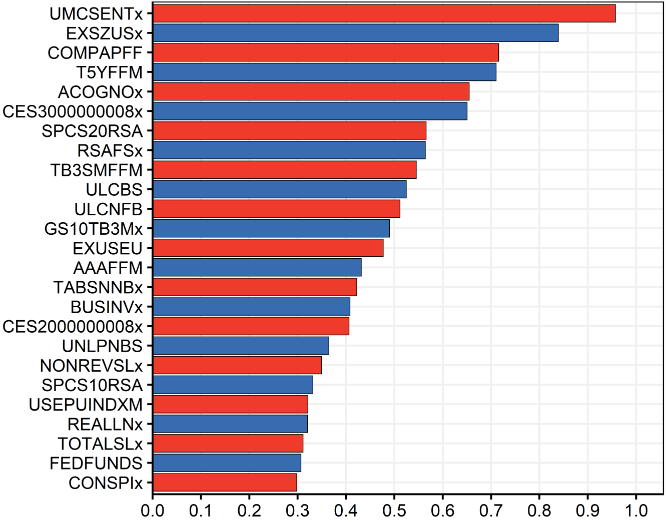
Adding a Volatility Hemisphere



An Extension and a Robustness Check



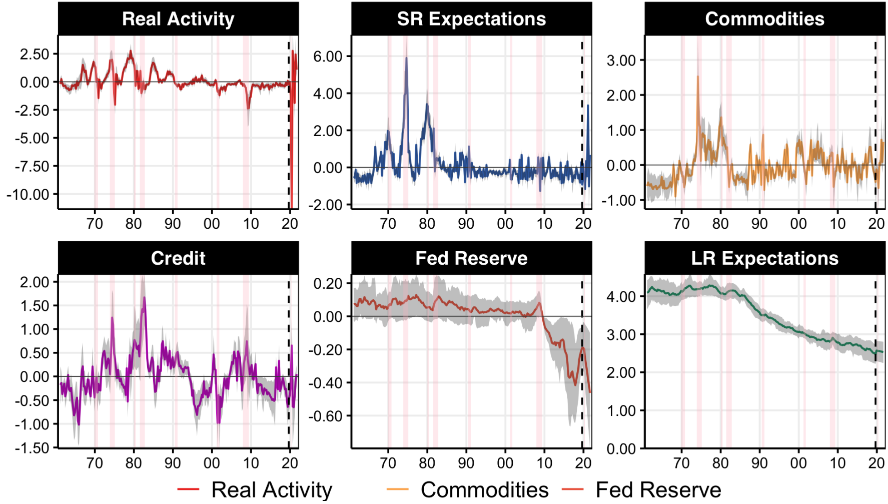
What is in the Sink?



- Mostly all forward looking variables that can characterize expectations about future economic outcomes

HNN-F-4NK

An empirical test of (Sims and Wu, 2019)'s 4 equations NK model



Using Traditional Proxies Won't Do.

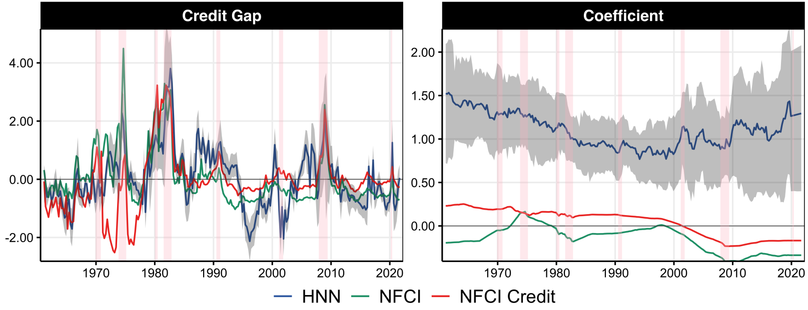


Figure: HNN-F-4NK's "credit conditions" and associated time-varying coefficient. Notes: Dashed line is the beginning of the out-of-sample. NBER recessions are in pink shadowing.

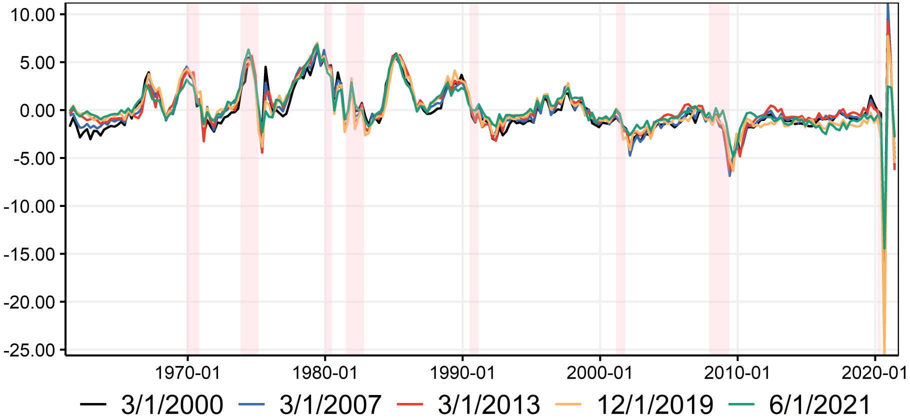
Conclusion

- Can use DL to extract a data-driven "output gap"
- Can use DL to think about the Phillips curve
- Can use DL to construct economically interpretable forecasts
- The HNN approach is widely applicable to models where the link between "theoretical variables" and "Excel variables" is sometimes muddy (neutral rate, taylor rules, term premium, etc).

Appendix

Is History Being Rewritten? (II)

Gap contributions ($h_{t,g}$) estimated with different samples



Variable Importance Procedure for $h_{t,j}$

- VI_k measure for a variable $X_{t,k}$ for $k \in \mathcal{H}_g$ works by shuffling randomly variable k (and all its attached transformations, i.e., lags and MARXs), recomputing **(but not re-estimating)** the prediction $h_j(\tilde{X}_t)$, and then comparing it to real one $h_j(X_t)$.

- Thus, the standardized VI_k, in terms of % of increase in MSE, is

$$VI_k = 100 \times \left(\frac{\sum_{t=1}^{T}(h_j(\tilde{X}_t) - h_j(X_t))^2}{Var(h_j(X_t))} - 1 \right)$$

- Intuitively, randomizing important variables will push $h_{j,t}$ far from its original estimate.

Coefficients of the Other Two Components