Uncovering Nonlinearities with Regression Anatomy: Online Appendix

Paul Bousquet

February 20, 2025

B.1 Full Expansion of FWL

Recall the general form from Section 2 of the paper

$$\omega_i = \frac{\text{cov}(1_{a\leq\varepsilon_t}, X_i^{\perp})}{\text{var}(X_i^{\perp})}$$

where $X_i^{\perp}$ is the residual from regressing X_i on the other elements in $\boldsymbol{X}_t$. We can unpack this definition to get things soley in terms of covariances and variance of terms of $\boldsymbol{X}_t$, which amounts to an expansion of the FWL theorem. To my knowledge, this expansion has not been done previously and for good reason – the full form amounts to several messy recursions that offer absolutely no insight to write out. However to motivate the use of deep learning to address one of the central issues in this paper, it may be useful to see why it's difficult to conjure up functional forms that will produce appropriate weighting.

For what follows, consider $\boldsymbol{X}$ to be a generic matrix of N covariates in a regression (which can include a vector of 1s) and X_i to be its i-th element. Keeping with notation from earlier, $X_i^{\perp}$ is the residual from X_i on the remaining elements of $\boldsymbol{X}$. WLOG, we will initially look at an example where $i = 1$. Further consider $X_n^{\perp_1}$ to be regressing the n-th element of $\boldsymbol{X}$ on its the remaining parts excluding X_1. Then

$$X_1^{\perp} = X_1 - \sum_{n=2}^{N} \frac{\text{cov}(X_1, X_n^{\perp_1})}{\text{var}(X_n^{\perp_1})} X_n$$

We can keep unpacking these terms but it should be clear that indexing is quickly going to become a nightmare because the "exclusions" will not be in a consistent ordering across the components (and sub-components, and sub-sub-components,...) of this summation. Things would have already got a bit messy notation wise had we done a formula for a generic $X_i^{\perp}$. So we will have to break this up into several parts.

First, we will deal with the covariance terms and keep the variance terms fixed. Again using $i = 1$ for indexing coherence and only focusing on the first term term in the sum above ($n = 2$), if we unpack a bit more we will have

$$\frac{\text{cov}(X_1, X_n^{\perp_1})}{\text{var}(X_n^{\perp_1})} = \frac{\text{cov}(X_1, X_2) - \frac{\text{cov}(X_2, X_3^{\perp_{1:2}})}{\text{var}(X_3^{\perp_{1:2}})}\text{cov}(X_1, X_3) + \ldots}{\text{var}(X_2^{\perp_1})}$$

where $X_3^{\perp_{1:2}}$ are the residuals of regressing X_3 on the remaining elements of $\boldsymbol{X}$ excluding X_1, X_2. To even coherently define the remaining terms in the numerator, new notation has to be introduced to deal with the order in which variables are excluded from the "sub-regressions". To address this, we will define things in chunks. Again keeping with the $i = 1$ and $n = 2$ case because it's the cleanest, note that

$$\frac{\text{cov}(X_1, X_2^{\perp_1})}{\text{var}(X_2^{\perp_1})} = \frac{C_{1,2} - C_{1,3}\left(\frac{C_{2,3}}{V_{2,3}} - C_{2,4}\left(\frac{C_{3,4}}{V_{3,4}V_{2,3}} - C_{3,5}(\dots) - \dots\right) - \dots\right) + \dots}{V_{1,2}}$$

where $C_{p,q}$ represents the covariance between elements p and q of $\boldsymbol{X}$ and $V_{p,q}$ is $\text{var}(X_q^{\perp_{1:p}})$. Ignoring the unavoidably ugly denoting of what terms correspond to, we can see a light of coherency at the end of the recursion tunnel. While there may now be a decipherable pattern to latch onto, this hasn't made the cases that are not $i = 1$ and $n = 2$ any less difficult to notate. So we will define a function relative to indexing. First observe

$$C_{1,3}\left(\frac{C_{2,3}}{V_{2,3}} - C_{2,4}\left(\frac{C_{3,4}}{V_{3,4}V_{2,3}} - C_{3,5}(\dots)\right)\right) = \sum_{k=2}^{N}(-1)^k \frac{C_{k,k+1}\prod_{j=1}^{k-1} C_{j,j+2}}{\prod_{j=2}^{k} V_{j,j+1}}$$

Everything else follows this structure, conditional on indexing. We can make an (ugly) generalization as follows. Let A be generic rearrangement of $\boldsymbol{X}$; i.e., $\boldsymbol{X}$ has indexing $\{1, \dots, N\}$ and A's indexing can be any permutation of this order. Define $I_k^{\{A\}}$ as the index of $\boldsymbol{X}$ corresponding to the k-th element of A (formally: a mapping $I(k; \{A\})$). Now define

$$S(\{A\}) = \sum_{k=2}^{N}(-1)^k \frac{C_{I_k^{\{A\}}, I_{k+1}^{\{A\}}} \prod_{j=1}^{k-1} C_{I_j^{\{A\}},‘}}{\prod_{j=2}^{k} V_{j,j+1}^{\{A\}}} \quad \text{with } V_{j,j+1}^{\{A\}} = \text{var}\left(X_{I_{j+1}^{\{A\}}}^{\perp_{I(1:j;\{A\})}}\right)$$

This will allow for a crawl towards completeness, burying as much of the index stumbling blocks as possible. Let P_N denote all permutations of $\{1, 2, \dots, N\}$. Define $P_N^{i,n} \subseteq P_N$ as permutations with i, n as the first elements:

$$P_N^{i,n} = \left\{\sigma \in P_N : \sigma(1) = i \;\&\; \sigma(2) = n\right\}$$

Then we can write the earlier expression $\text{cov}(X_1, X_2^{\perp_1})$ in the general case as

$$\text{cov}(X_i, X_n^{\perp_i}) = C_{i,n} - \boldsymbol{\Sigma}_{i,n} \quad \text{with } \boldsymbol{\Sigma}_{i,n} = \sum_{\sigma \in P_N^{i,n}} S(\{\sigma\})$$

At the very beginning we started with

$$\omega_i = \frac{\text{cov}(1_{a \leq \varepsilon_t}, X_i^{\perp})}{\text{var}(X_i^{\perp})}$$

And now we can write $\text{cov}(1_{a \leq \varepsilon_t}, X_i^{\perp})$ compactly as

$$\text{cov}(1_{a \leq \varepsilon_t}, X_i^{\perp}) = \text{cov}(1_{a \leq \varepsilon_t}, X_i) - \sum_{n \geq 1: n \neq i}^{N} \text{cov}(1_{a \leq \varepsilon_t}, X_n) \cdot \left(\frac{C_{i,n} - \boldsymbol{\Sigma}_{i,n}}{\text{var}(X_n^{\perp_i})}\right)$$

We are not out of the woods yet because we skipped unpacking the variance terms. But once that has been done we will have finished "simplifying", in that arbitrarily complex regressions can be defined in terms of estimands that feature only explicit variance and covariance terms.[1]

The strategy to deal with the variance terms will be very similar and hopefully easier to digest now that we have some machinery to work with. Again to deal with the simplest case ($i = 1$) first,

$$\mathrm{var}(X_1^{\perp}) = \mathrm{var}(X_1) + \sum_{n=2}^{N}\sum_{m=2}^{N} \frac{\mathrm{cov}(X_1, X_n^{\perp_1})}{V_{1,n}} \frac{\mathrm{cov}(X_1, X_m^{\perp_1})}{V_{1,m}} C_{n,m} - \sum_{n=2}^{N} \frac{\mathrm{cov}(X_1, X_n^{\perp_1})}{V_{1,n}} 2C_{1,n}$$

using the notation as before for V and C. Once more, we have a situation where everything will follow this pattern, less indexing. The first layer is simple to write

$$\mathrm{var}(X_i^{\perp}) = \mathrm{var}(X_i) + \sum_{m,n\geq 1: m,n\neq i}^{N}\sum^{N} \frac{\boldsymbol{\Sigma}_{i,n}}{V_{1,n}} \frac{\boldsymbol{\Sigma}_{i,m}}{V_{1,m}} C_{n,m} - \sum_{n\geq 1: n\neq i}^{N} \frac{\boldsymbol{\Sigma}_{i,n}}{V_{1,n}} 2C_{i,n}$$

The only thing remaining is to expand this definition so that it holds as terms are continually added to $\perp$ in $X_i^{\perp}$ (i.e., in the FWL regressions, some terms have already been partialed out and won't be included). To do this, we need to make the indexing of $\boldsymbol{\Sigma}_{i,n}$ a bit more flexible. Define

$$SV(i;\{O\}) = \mathrm{var}(X_i^{\perp_{\{O\}}}) = \mathrm{var}(X_i) + \sum_{m,n\geq 1: m,n\notin O}^{N}\sum^{N} \frac{\boldsymbol{\Sigma}^c_{\{O\},n}}{SV(n;\{O,i\})} \frac{\boldsymbol{\Sigma}^c_{\{O\},m}}{SV(m;\{O,i\})} C_{n,m} - \sum_{n\geq 1: n\notin O}^{N} \frac{\boldsymbol{\Sigma}^c_{\{O\},n}}{SV(n;\{O,i\}} 2C_{i,n}$$

where O is a set of unique integers $o \in [1,N]\backslash\{i\}$ and

$$\boldsymbol{\Sigma}^c_{\{O\},n} = \sum_{\sigma\in P_-^{\{O\}}} S(\{\sigma\})$$

$$\text{with } \; P_-^{\{O\}} = \left\{\sigma \in \mathbb{Z}^{N-|O|} : \sigma \cup \{O\} \in P_N \;\&\; \forall n, \nexists m \text{ s.t } \sigma(m) = O(n) \right\}.$$

Noting that for any $\sigma \in P_N$, $SV\Big(i, \big\{\sigma\backslash\{i\}\big\}\Big) = \mathrm{var}(X_i)$, our nightmare is finally over.

B.2 Deep Learning Implementation

Work is still ongoing to fine tune the algorithm and therefore to sharpen these recommendations. Also, the Online Appendix from this point forward is a work in progress, but for the remaining sections its a matter of compilation, rather than work that remains to be completed.

One consideration is complexity. I've found that convergence occurs rapidly, but convergence may not be to a collection of satisfactory functions (i.e., the neural net essentially gets stuck at a local minima). So while this may be more feasible without a GPU, the iterative nature of refining the loss computation may make this less feasible without access to hardware designed for efficient deep learning training. I hope to eventually share functions on this paper's GitHub repository that work well for standard normal shocks. When applied to other shock series, the performance will not be perfect but may be acceptable.

[1] Of course, $(X'X)^{-1}X'Y$ is a better simplification under any sensible definition. "disambiguating" may be more appropriate

B.3 Standard Errors for Generated Regressors

Recall the generated regressor functions defined in Section 3 $\{f_i\}_{i=1}^4$. For clarity, we are interested in functions of a shock ε_t that is continuously distributed on $a \in I \subset \mathbb{R}$. Each of this function has a designated "peak" c_i and a set I_i with endpoints left$_i$ and right$_i$). These functions are defined in terms of the empirical CDF $F_N(\cdot)$.

Specifically, for each $a \in I$

$$f_i(a) = \begin{cases} 0 & a \notin [\text{left}_i, \text{right}_i) \\ -[F_N(c_i) - F_N(\text{left}_i)]^{-1} & a \in [\text{left}_i, c) \\ [F_N(\text{right}_i) - F_N(c_i)]^{-1} & a \in (c, \text{right}_i) \end{cases}$$

with slight abuse of notation if left$_i = -\infty$. In a regression of y on $\{f_i\}_{i=1}^4$, the estimands will be defined in terms of the CDF $F(\cdot)$. Define p_{iL} as $F(c_i) - F(\text{left}_i)$ and $p_{iR} = F(\text{right}_i) - F_N(c_i)$. The estimand β_i on f_i is

$$\beta_i = \frac{\text{cov}(y, f_i)}{\text{Var}(f_i)} = \frac{\bar{y}_{iR} - \bar{y}_{iL}}{\frac{1}{p_{iL}} + \frac{1}{p_{iR}}},$$

where $\bar{y}_{iL}$ and $\bar{y}_{iR}$ are the means of y on the subsets of I_i given by p_{iL} and p_{iR}. To see this, recall f_i is mean 0. So

$$\text{cov}(y, f_i) = \mathbb{E}[y f_i] = -\frac{1}{p_{iL}} \mathbb{E}[y \cdot \mathbb{1}_{[\text{left}_i, c_i)}] + \frac{1}{p_{iR}} \mathbb{E}[y \cdot \mathbb{1}_{[c_i, \text{right}_i)}] = -\frac{1}{p_{iL}} \bar{y}_{iL} \cdot p_{iL} + \frac{1}{p_{iR}} \bar{y}_{iR} \cdot p_{iR} = \bar{y}_{iR} - \bar{y}_{iL}$$

Because this estimand is formed with respect to a generated regressor, we need to adjust the standard errors. Adjustment is done using the delta method. Differentiating

$$\frac{\partial \beta_i}{\partial p_{iL}} = \beta_i \cdot \frac{p_{iR}}{p_{iL}(p_{iL} + p_{iR})} \quad \underline{\textbf{and}} \quad \frac{\partial \beta_i}{\partial p_{iR}} = -\beta_i \cdot \frac{p_{iL}}{p_{iR}(p_{iL} + p_{iR})}$$

The adjustment takes the form of [2]

$$\left(\frac{\partial \beta_i}{\partial p_{iL}}\right)^2 \text{Var}(p_{iL}) + \left(\frac{\partial \beta_i}{\partial p_{iR}}\right)^2 \text{Var}(p_{iR}).$$

Using sample analogs, standard errors are the square soot of the sum of the usual Huber-White variance and

$$\frac{\hat{\beta}_i^2}{N} \left(\frac{\hat{p}_{iR}(1 - \hat{p}_{iL})}{\hat{p}_{iL}(\hat{p}_{iL} + \hat{p}_{iR})^2} + \frac{\hat{p}_{iL}(1 - \hat{p}_{iR})}{\hat{p}_{iR}(\hat{p}_{iL} + \hat{p}_{iR})^2} \right)$$

where $\widehat{\text{Var}}(p_{iL}) = \frac{\hat{p}_{iL}(1-\hat{p}_{iL})}{N}$ (and similar for p_{iR}).

B.4 Details for Empirical Application in Section 4.3

Following Ramey (2016), outcome variables are the Consumer Price Index (CPI), industrial production, 1 year treasury yields, excess bond premium (Favara et al., 2016), unemployment, and add real consumption expenditures (all monthly frequency).[3] Control variables also include lagged interest rates, monetary policy uncertainty (Husted

[2]Note that f_i is not differentiable at c_i

[3]Some of these variables are highly non-stationary (McCracken and Ng, 2016). Montiel Olea and Plagborg-Møller (2021) show LP is remarkably robust to the presence of unit roots and non-stationary variables. I find some anecdotal support for this: estimating in differences and summing for the cumulative effect in levels produces very similar IRFs to estimating in levels directly

et al., 2020), an indicator for the ZLB, and a healthy number lags (12) of both outcome and controls following our discussion of standard errors. Data is sourced from FRED unless noted otherwise and the maximum sample periods are retained. Based on the availability of the shock series used in the paper, this is ultimately not an issue, but in general pre-1983 target funds rate data should be discarded to reflect incongruities in Fed policy norms (Thornton, 2006; Aruoba and Drechsel, 2024). Related, to aggregate to changes as intended by policymakers, a few earlier instances of "intermediate" changes (e.g., adjust 12 basis points immediately and 25 more in a few weeks) are cumulated. This adds another reason to be concerned about temporal aggregation bias (Jacobson et al., 2024) and likely to discard results at early horizons. I focus on CPI and the joint picture of output painted by industrial production, consumption, and unemployment in order to take the findings directly to models. Outcome variables are cumulative log differences, yielding an approximate percent change interpretation: $\widehat{\alpha}_h$ is represents the percent change in levels h periods after a shock at t. At $h = 12$, this takes a nice form of year over year growth.

The LP framework described in the paper can be used to illustrate possible nonlinearities, which I refer to as size and size effects. The most straightforward way to think of these effects is as functions of parameters. For the simple case of plotting in levels, the objects of interest are

$$\text{Size Effect}_h : \hat{\alpha}_h^B - \hat{\alpha}_h^S \qquad \text{Sign Effect}_h : \hat{\alpha}_h^P + \hat{\alpha}_h^N$$

A size effect exists if we can conclude the difference in the big and small (regime) coefficients are distinct from 0 and a sign effect exists if positive and negative coefficients have different magnitudes. This is complicated slightly by wanting to allow for both types of non-linearities simultaneously: we want to see if a size effect exists for both cuts and hikes and a sign effect exists for both big and small changes, in other words 4 graphs per outcome variable. Grouping is thus by type of nonlinearity, rather than outcome.

For sign effects, we can interpret positive statistically significant results as evidence for a "pushing on a string" narrative, the idea that it's more difficult (especially in recessions) for expansionary monetary policy to stimulate the economy than it is for contractionary policy to suppress it. Even if the individual point estimates go against what standard theory might suggest, we abstract from the notion of puzzles by simply focusing on one estimate relative to the other. For instance, if the coefficient of a big cut on CPI is -3 and the coefficient for a big hike is 2, these estimates are consistent with the string story because the the big hike's contractionary power, albeit a lack of one, is still greater than the expansionary effect of a big cut.

Before gauging how models holds up to the data-based findings presented in the main body of the paper, it's important to have a sense of what, if anything, can make these results weaken when pushed. Changing the lag order, adding and removing controls, estimating in differences vs. levels, different measures for inflation, bias-correcting point estimates (Herbst and Johannsen, 2024) and using LP instead of LP-IV in general do not yield IRFs with different interpretations, even under various combinations of these factors. One area where there is some sensitivity is sample selection and size, which is especially not surprising given the zero lower bound period. One option would be using a non-linear filtering procedure (e.g., Farmer, 2021) to construct a shadow interest rate, or a measure of how interest rates "would have moved" if the ZLB didn't bind.

A more involved critique of model-free estimation is an inability to account for state-dependence. For example, many past efforts try to allow for responses in boom and bust cycles to be asymmetric. With respect to interest rate shocks of a given size, another worry could be that beliefs about the future path of policy may not be updated in the same for different histories of action. The econometric concern is that these local projection coefficients amount to weighted averages and these weights could be biased if the joint distribution of the shock and state space has disparate behavior from a product of their marginals. In a regression context, this essentially amounts to the difference between including a variable as a control and additionally interacting it with the shock. The work of Rambachan and Shephard (2021), Kolesár and Plagborg-Møller (2024), and results in the main body of the paper show that these estimates average out under arbitrary nonlinearities, so within this setting there is less room for concern. But because of the limited sample size, it's worth taking note of other, more directed approaches. Jordà (2023), following the revelation of Gonçalves et al. (2024) that the previous default methodology can severely distort impulse responses, provides a framework incorporating interaction terms to estimate state-dependent effects. Gonçalves et al. (2024) themselves suggest non-parametric estimation, which has an over-parameterization problem with or without instrumenting (i.e., in either case, control variables must be shed).

B.5 Details for Quantitative Application in Section 4.3

Description	Equation	#
Consumption Euler Equation	$1=\beta\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\tau}\frac{R_t}{\Pi_{t+1}\tilde{A}_{t+1}}\right]$	(1)
Definition for Real Wages	$\Delta_t^w=\frac{W_t}{W_{t-1}}\cdot\tilde{A}_t$	(2)
Resource Constraint	$\frac{G_t-1}{G_t}\cdot Y_t+C_t=Y_t(1-\Phi_t^p)-W_tY_t\cdot\Phi_t^w$	(3)
Wage Equation, Household's problem	$\frac{\chi_h}{\lambda_w W_t}C_t^{\tau}Y_t^{\frac{1}{\nu}}+(1-\Phi_t^w)\left(1-\frac{1}{\lambda_w}\right)=$	(4)
	$\Delta_t^{w_{nom}}\cdot\Phi_t^{\prime w}-\beta\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\tau}\frac{\Pi_{t+1}(\Delta_{t+1}^w)^2}{\tilde{A}_{t+1}}Y_{t+1}\cdot\Phi_{t+1}^{\prime w}\right]$	
Price Equation, Intermediate Firms problem	$(1-\Phi_t^p)+\beta\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\tau}\Phi_{t+1}^{\prime p}\Pi_{t+1}Y_{t+1}\right]=\frac{\mu_t}{\Lambda_t}+\Phi_t^{\prime p}\Pi_t$	(5)
Hours Equation	$W_t=(1-\Phi_t^p)-\mu_t$	(6)
Adjustment Costs, Nominal Wages	$\Phi_t^w=\frac{\phi_w}{\psi_w^2}\left(e^{-\psi_w(\Delta_t^{w_{nom}}-\gamma\pi^*)}+\psi_w(\Delta_t^{w_{nom}}-\gamma\pi^*)-1\right)$	(7)
Adjustment Costs, Prices	$\Phi_t^p=\frac{\phi_p}{\psi_p^2}\left(e^{-\psi_p(\Pi_t-\pi^*)}+\psi_p(\Pi_t-\pi^*)-1\right)$	(8)
Derivative, Adjustment Costs Nominal Wages	$\Phi_t^{\prime w}=\frac{\phi_w}{\psi_w}\left(1-e^{-\psi_p\left(\Delta_t^{w_{nom}}-\gamma\pi^*\right)}\right)$	(9)
Derivative, Adjustment Costs to Prices	$\Phi_t^{\prime p}=\frac{\phi_p}{\psi_p}\left(1-e^{-\psi_p\left(\Pi_t-\pi^*\right)}\right)$	(10)
Taylor Rule	$R_t=\exp(r_t);\quad r_t=\rho_r r_{t-1}+(1-\rho_r)r_t^*+\sigma_r\varepsilon_r$	(11)
TFP Growth	$\tilde{A}_t=\exp(a_t);\quad a_t=(1-\rho_a)\log\gamma+\rho_a a_{t-1}+\sigma_a\varepsilon_a$	(12)
Government Spending Shocks	$G_t=\exp(g_t);\quad g_t=(1-\rho_g)\log g^*+\rho_g g_{t-1}+\sigma_a\varepsilon_g$	(13)
Price Markup Shock	$\Lambda_t=\exp(\lambda_t);\quad \lambda_t=(1-\rho_p)\log\lambda_{p_{ss}}+\rho_p\lambda_{t-1}+\sigma_p\varepsilon_p$	(14)
Output change	$\Delta_t^y=Y_t/Y_{t-1}$	(15)
Nominal wage change	$\Delta_t^{w_{nom}}=\Delta_t^w\Pi_t$	(16)
Interest rate target	$r_t^*=\log\left(\frac{\gamma}{\beta}\cdot\pi^*\right)+\psi_1\log\left(\Pi_t/\pi^*\right)+\psi_2\left(a_t+\log\left(\Delta_t^y/\gamma\right)\right)$	(17)

- For the set of draws that came out of our Metropolis- Hastings routine, I simulated data of 400 observations for each group of parameters to align with the US data sample size. Analogous control variables are included (lagged interest rates, zero lower bound, unemployment, output and interest rate variance) and plots are in terms of standard deviations to abstract away from any differences between model-simulated and US data. This is described and justified further in the next subsection.

- The priors are largely from Aruoba et al. (2017). Because of the difference in sample period, I scaled down the priors for annualized output growth (μ_y) and inflation (μ_π), as well as β^{-1}. In fact, it's actually not possible for this model to generate a steady state that matches the data. Steady state interest rates are $\mu_y+\mu_\pi+400(\beta^{-1}-1)$. If μ_y and μ_π are picked to match inflation data, you must pick $\beta>1$ to match interest rate data.

- Related^, the authors start the M-H algorithm at the mode of the linear model and then manually append starting values for the asymmetry parameters and fix the diagonals of the inverse hessian at 4. Asymptotically, there's nothing wrong with this, but I'm going to play around with this to see how sensitive the algorithm is to initial values and priors. From what I've done so far, it seems like there's sensitivity on both fronts. One key point is that when estimating the linear mode, they include ψ_w, which is not identified in a first order estimation.

- A running "diary" of some things I've discovered working with this model/codes can be found here. Ongoing work is paying close attention to the particle filter implementation

- For consistency in the comparative static illustrations, it was necessary to make sure this mode line was the same across plots, but that meant the same series of shocks would need to be used for all sets of simulated data, which could paint a misrepresentative picture for a short sample size. So I plotted the median estimate for 100 samples for each parameter group (for simulation i, seed was set to $i = \{1, \ldots, 100\}$).

  – Ideally, this would be done for the Bayesian IRFs, but that would take a month to run. Implementation needed to be extremely parsimonious – because each loop of the LP file performs 25*number of outcome variables calls to regress, I randomly selected 10,000 draws of the post burn-in M-H output.

	h		
	0	1	2
Big Cut	-18.8%	-5.5%	-1.7%
Big Hike	36.6%	5.5%	0.4%

Table 1: Average % Deviation from i^*, h periods after large change in i_t

Next, I show that the model is capable of generating any type of nonlinearity on impact, but the effects dissipate quickly. To make efficient use of space, the exact figures are relegated to the very end of the Appendix, but there are hyperlinks to each. Again, what we learned from this exercise is that any sort of nonlinearity desired can be generated on impact using the right combination of asymmetry parameters, but it does not last for even one period longer in most cases.

Size Effects

	Description	Anything Interesting? (all at $h = 0$)	Link
1	$\psi_p \uparrow$	(slightly) amplifies negative size effect for hikes on π at $h = 0$	Figure 1
2	$\psi_p \downarrow$	(slightly) amplifies all $h = 0$ size effects except for hike on π	Figure 2
3	$\psi_w \uparrow$	(slightly) increases the positive size effect for cuts on Y at $h = 0$	Figure 3
4	$\psi_w \downarrow$	No.	Figure 4
5	$\psi_p \uparrow$, $\psi_w \uparrow$	amplifies size effect (-) of cuts on Y, depresses size effect (+) of hikes on Y	Figure 5
6	$\psi_p \downarrow$,$\psi_w \downarrow$	(slightly) amplifies negative size effect for hikes on π at $h = 0$	Figure 6

Sign Effects

	Description	Anything Interesting? (all at $h = 0$)	Link
1	$\psi_p \uparrow$	depressed all $h = 0$ sign effects except for small changes on π	Figure 7
2	$\psi_p \downarrow$	low values reversed the direction of the sign effect for big changes on π.	Figure 8
3	$\psi_w \uparrow$	Depresses small change on Y size effect and amplifies everything else	Figure 9
4	$\psi_w \downarrow$	(slightly) amplified sign effect of big changes on π and small changes on Y	Figure 10
5	$\psi_p \uparrow$, $\psi_w \uparrow$	depressed sign effect of small changes on Y and amplified everything else	Figure 11
6	$\psi_p \downarrow$,$\psi_w \downarrow$	reversed the direction of sign effect for big changes on π	Figure 12

B.6 Estimation in Terms of Standard Deviations

In a linear regression, coefficients are the estimated effect of a marginal (size), positive (sign) change. If we normalize our previous definitions by the standard deviation of the coefficient corresponding to this linear "default", we have an alternative formulation of size and sign effects in terms of standard deviations instead of percent change in levels at a given horizon. For example, if $\frac{\hat{\alpha}_{BC} - \hat{\alpha}_{SC}}{\sigma_{SC}} = 3$, the interpretation is that the big cut coefficient amounts to a 3 standard deviations away realization of the small cut coefficient. Additional intuition can be gleaned by noticing that if we instead normalized by the standard deviation of the entire (original) definition, we would simply have a t-statistic. This approach has the advantage of the y-axis having a uniform representation across all outcomes of interest and arguably removes some of the subjectivity implicit in deciding what % constitutes a meaningful effect for a given variable-horizon combination. Put differently, this representation sends a similar signal to the results of a hypothesis test (is there enough evidence from data to infer these parameters are drawn from distinct distributions), but unlike a t-statistic the units lend themselves more to economic meaning (moment of the distribution for our baseline coefficient, rather than a general normal distribution). Another motivation is model comparison. In principle, percent change is a "unitless" point of comparison. But in a small sample setting, having parameters in a DSGE model that dictate growth rates can induce distortions in scaling and correlations relative to time series data that add meaningless noise to analogous estimations of our objects of interest. By using standard deviations, everything is normalized by whatever scale persists in the DGP, meaning the finite sample idiosyncrasies are softened. Ultimately, percent change in levels carries more real world weight, but standard deviations add indispensable context.

References

Aruoba, S. Boragan, and Thomas Drechsel. 2024. “Identifying Monetary Policy Shocks: A Natural Language Approach.” Working paper.

Aruoba, S. Borağan, Luigi Bocola, and Frank Schorfheide. 2017. “Assessing DSGE model nonlinearities.” *Journal of Economic Dynamics and Control*, 83 34–54.

Farmer, Leland E. 2021. “The discretization filter: A simple way to estimate nonlinear state space models.” *Quantitative Economics*, 12(1): 1–45.

Favara, Giovanni, Simon Gilchrist, Kurt F. Lewis, and Egon Zakrajšek. 2016. “Updating the Recession Risk and the Excess Bond Premium.”

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. 2024. “State-dependent local projections.”

Herbst, Edward P., and Benjamin K. Johannsen. 2024. “Bias in local projections.” *Journal of Econometrics*, 240(1): .

Husted, Lucas, John Rogers, and Bo Sun. 2020. “Monetary Policy Uncertainty.” *Journal of Monetary Economics*, 115 20–36.

Jacobson, Margaret M., Christian Matthes, and Todd B. Walker. 2024. “Temporal Aggregation Bias and Monetary Policy Transmission.” March, Working Paper.

Jordà, Òscar. 2023. “Local Projections for Applied Economics.”

Kolesár, Michal, and Mikkel Plagborg-Møller. 2024. “Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly.” Working Paper.

McCracken, Michael W., and Serena Ng. 2016. “FRED-MD: A Monthly Database for Macroeconomic Research.” *Journal of Business & Economic Statistics*, 34(4): .

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. “Local Projection Inference is Simpler and More Robust Than You Think.” *Econometrica*, 89(4): 1789–1823.

Rambachan, Ashesh, and Neil Shephard. 2021. “When do common time series estimands have nonparametric causal meaning?” 10, Working paper.

Ramey, Valerie. 2016. “Macroeconomic Shocks and Their Propagation.” In *Handbook of Macroeconomics*. Chap. 2.

Thornton, Daniel L. 2006. “When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us.” *Journal of Money, Credit, and Banking*, 38(8): 2039–2071.

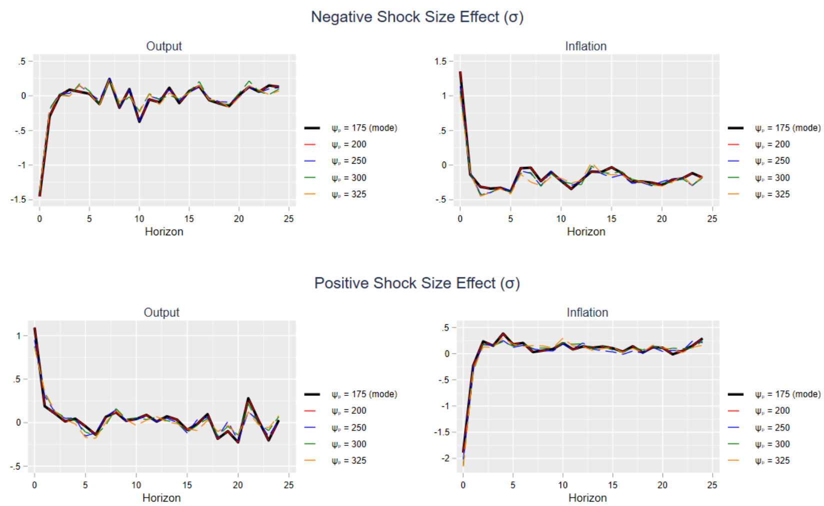

Figure 1: (click to go back to tables)

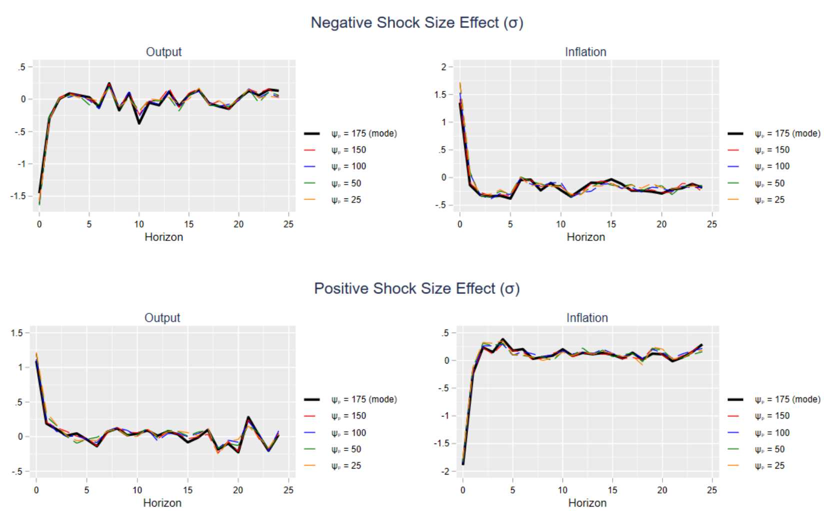

Figure 2: (click to go back to tables)

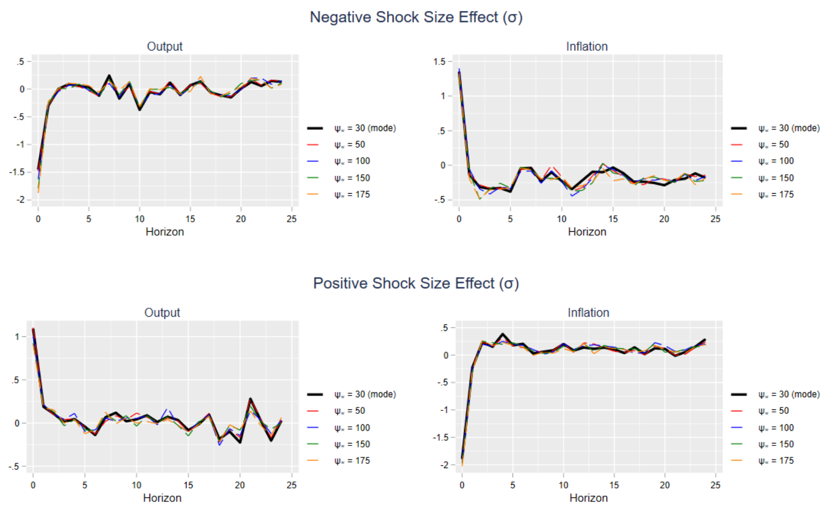

Figure 3: (click to go back to tables)

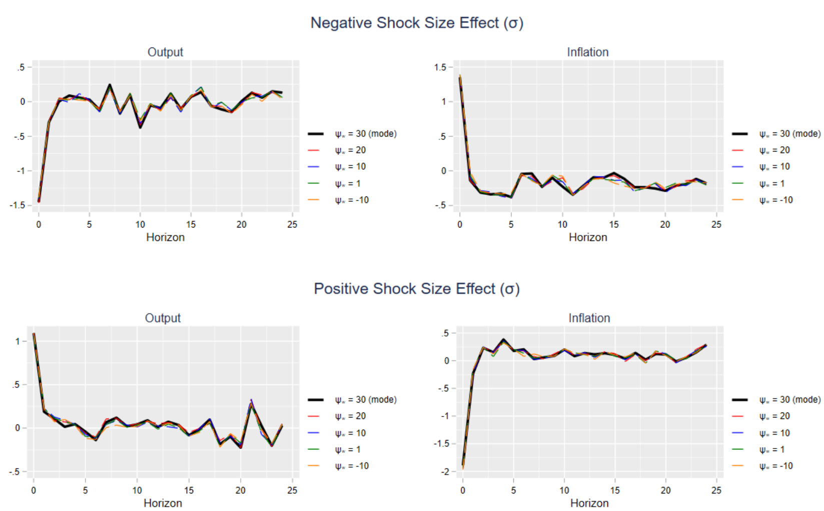

Figure 4: (click to go back to tables)

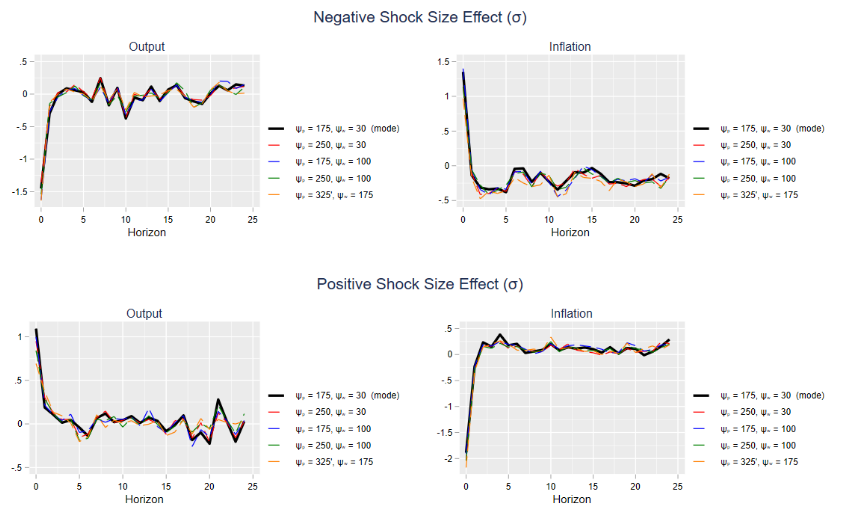

Figure 5: (click to go back to tables)

Figure 6: (click to go back to tables)

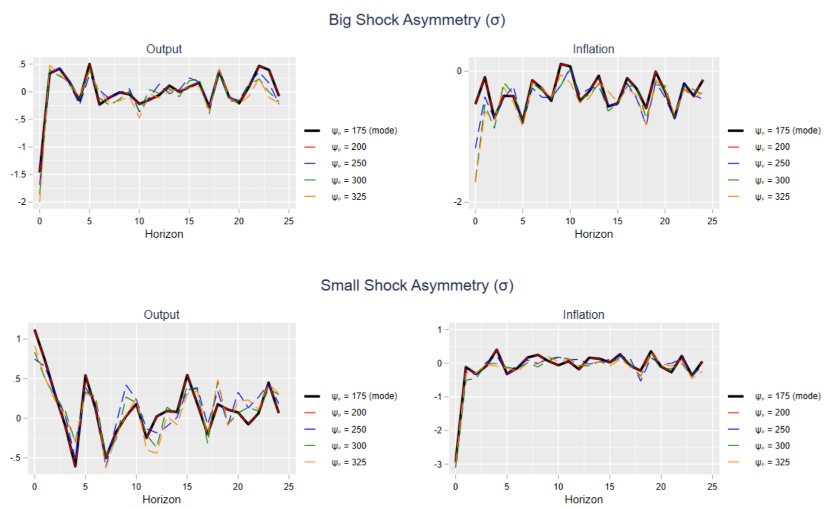

Figure 7: (click to go back to tables)

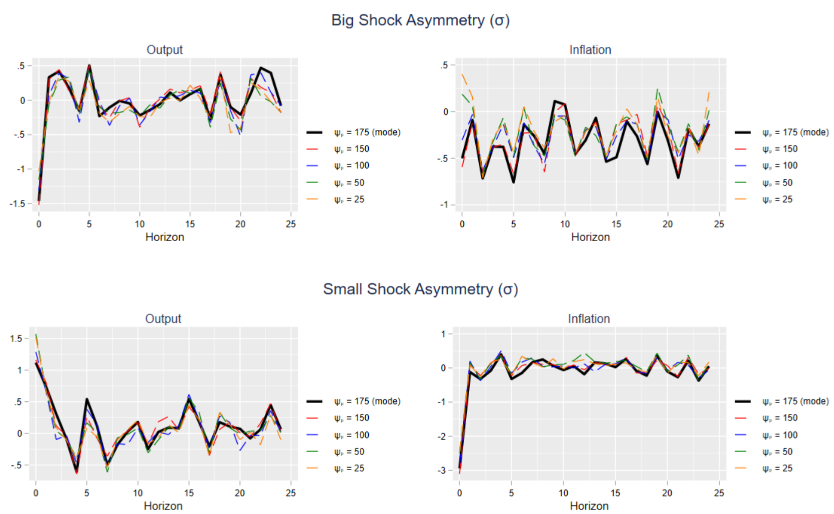

Figure 8: (click to go back to tables)

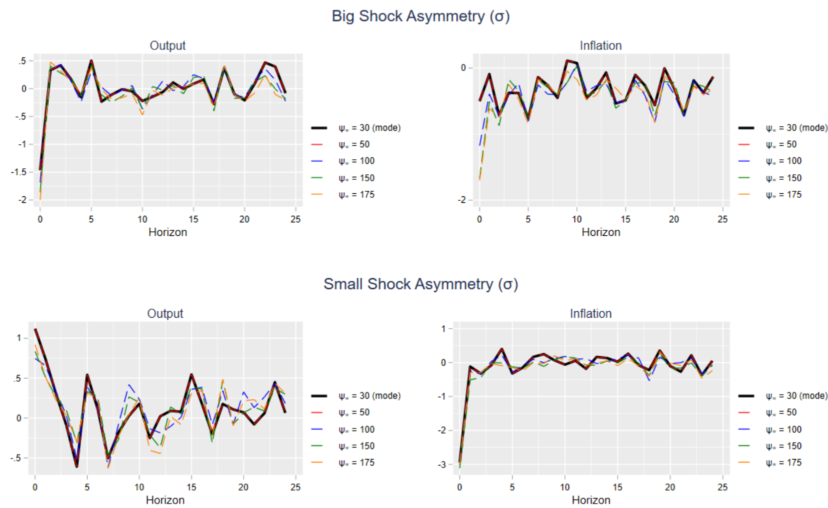

Figure 9: (click to go back to tables)

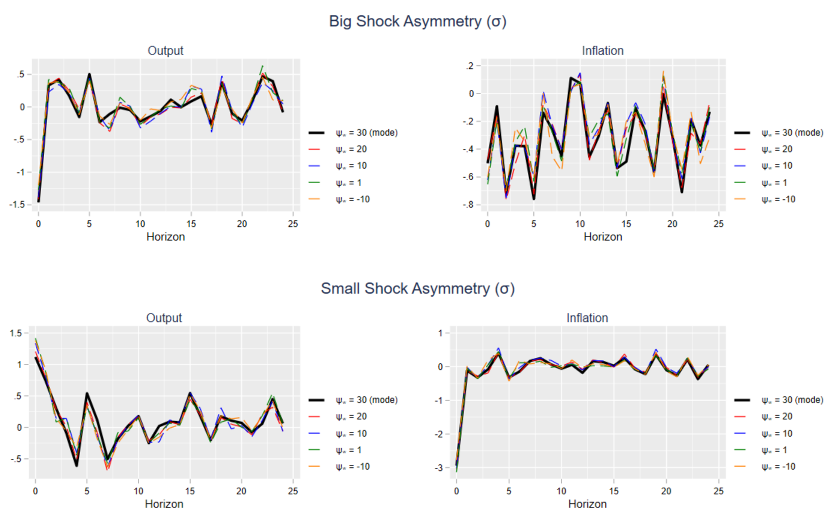

Figure 10: (click to go back to tables)

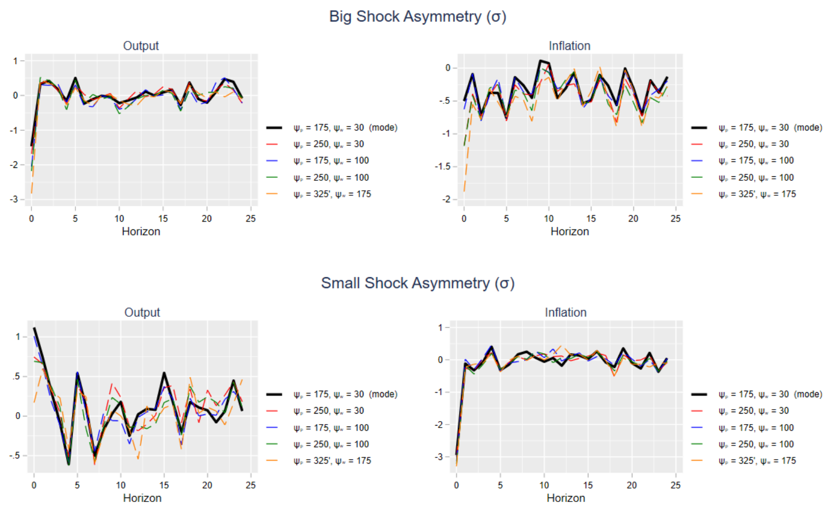

Figure 11: (click to go back to tables)

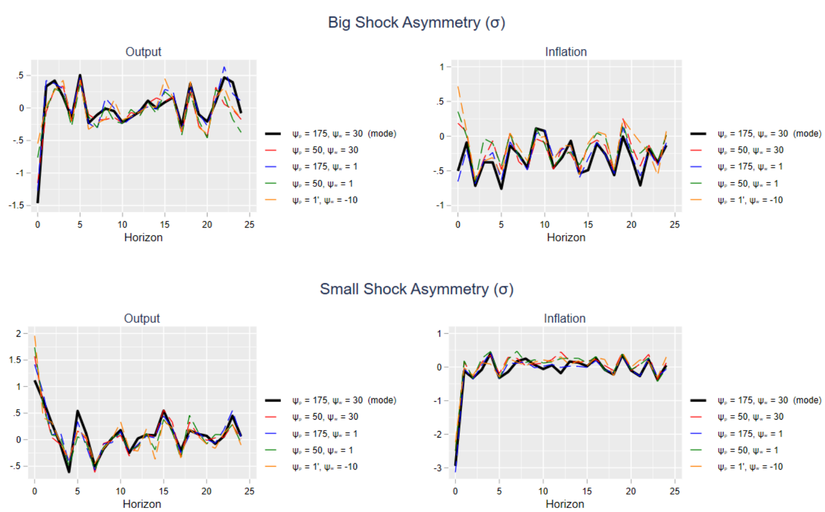

Figure 12: (click to go back to tables)