Federal Reserve Bank of Cleveland Working Paper Series

Online Appendix for Post-COVID Inflation Dynamics: Higher for Longer

Randal J. Verbrugge and Saeed Zaman

Working Paper No. 23-06R

June 2023

Suggested citation: Verbrugge, Randal J. and Saeed Zaman. 2023. "Online Appendix for Post-COVID Inflation Dynamics: Higher for Longer." Working Paper No. 23-06R. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202306r.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Online Appendix for Post-COVID Inflation Dynamics: Higher for Longer

Randal Verbrugge
Federal Reserve Bank of Cleveland and NBER/CRIW, 1455 E. 6th St., Cleveland, OH 44114
Email: randal.verbrugge@clev.frb.org.

Saeed Zaman
Federal Reserve Bank of Cleveland, 1455 E. 6th St., Cleveland, OH 44114
Email: saeed.zaman@clev.frb.org.

May 31, 2023
(First version: January 13, 2023)

Acknowledgments: We thank Rick Ashley, Marta Banbura, David Berger, Mark Bils, Elena Bobeica, Mark Bognanni, Andrew Figura, Ed Knotek, David Lebow, James Mitchell, Norm Morin, James Morley, Michael Palumbo, Matthias Paustian, Katia Peneva, Juan Rubio-Ramirez, Jeremy Rudd, Arsenios Skaperdas, Jim Stock, Brad Strum, and our other research colleagues at the Federal Reserve Bank of Cleveland, and seminar participants at the Board of Governors and the European Central Bank, for excellent comments that helped to greatly improve this paper. Any errors are our own. The views stated herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

Online Appendix

Appendix A: Partitioning the Jobless Unemployment Rate

A.1 Overview

To partition the jobless unemployment rate while applying the Ashley/Verbrugge (2022a,b) method (described below), we use the Iacobucci-Noullez (2005) filter, setting $k = 4$ (i.e., using 4 quarters of univariate forecasts in each rolling window). The Iacobucci-Noullez filter introduces no phase shift (unlike, e.g., the Christiano-Fitzgerald (2003) filter).[1] Following Ashley and Verbrugge (2023), we choose frequency cutoffs so that the jobless unemployment rate is partitioned into fluctuations lasting longer than 4 years (termed u_t^{lowgap}, for low-frequency gap), fluctuations lasting between 1 year and 4 years (termed $u_t^{medfreq}$, for medium-frequency), and transient fluctuations. Transient fluctuations were found to be unimportant drivers of inflation, and so were omitted. The two more persistent unemployment components are plotted in Figure 2. This figure also demonstrates how unusual the COVID collapse and recovery were. In particular, the low-frequency gap rose, notably much more sharply than usual. Meanwhile, the medium-frequency component also fell very sharply back to zero, dropping to historic lows.

The first step is to *partition* the real-time unemployment rate u_t into 3 persistence components – $u_t = \sum_{j=1}^{3} u_{j,t}$ – which *by construction* add up to the original series. These 3 persistence components partition the variation in u_t into monotonically decreasing levels of persistence or, equivalently, increasing frequency levels. These components are obtained from the sample data using a moving window (augmented with a k-quarter forecast) to filter the u_t data at each time t in a one-sided (backward-looking) manner. This approach mitigates end-of-sample filter distortions, ensures that parameter estimates are consistent, and retains both the causality structure of the data-generating process and any orthogonality conditions that are present in the unfiltered data. The Ashley/Verbrugge persistence-dependent regression methodology then merely replaces $u_t - u_t^*$ with these 3 persistence components, estimating a separate coefficient for each. (We note in passing that we subtract u_t^* from the most persistent component. That way, the 3 components add up to an unemployment gap, the typical Phillips curve specification in the literature.) Simulation evidence in Ashley, Tsang, and Verbrugge (2020) and Ashley and Verbrugge (2022b) indicate that the method yields reliable coefficient estimates and inferences, for both linear and nonlinear data-generating processes. Below we summarize why partitioning, one-sided filtering, augmentation or 'padding' with forecasts, and restriction of the filtering

[1] Use of the CF filter in the AV method produced qualitatively similar results, though it calls for the predicted recession to begin a quarter or two later. For a comparison of the use of different filters for frequency-dependent regression (as well as the sensitivity of results to forecast parameters), see Ashley and Verbrugge (2020b). Using the CF filter with the Ashley/Verbrugge method mitigates its phase shift in any case.

solely to the u_t data are all *essential* for obtaining reliable inferences.**A.2 One-Sided Filtering Method of Ashley and Verbrugge**

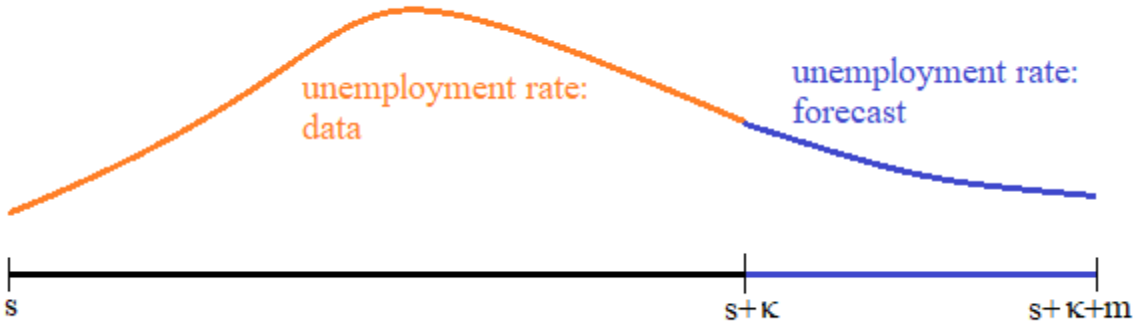


Figure A.1: One-sided filtering of unemployment rate data at date s+τ
(using a two-sided filter, from time s to time s+k+m)

A2.1 Description of one-sided filtering

In brief, one undertakes the one-sided filtering by running a window through the data. Over each window, one saves the decomposition at the final *data point* in the window. Then one increments the window by one quarter. However, each window includes not just data but also a second component that is a forecast. In other words, each window includes data augmented with a forecast.

To explain this in more detail, consider Figure A.1. We wish to compute the decomposition of the unemployment rate at time s+κ. As is well-known, obtaining the decomposition at s+κ by using a two-sided filter from time s to time s+κ would yield estimates with very poor properties. In particular, the resultant time series would (for most filters) incorporate a pronounced phase shift, in addition to being highly inaccurate; this inaccuracy is due to the well-known "edge effect" problem plaguing all filters.

Both the phase-shift and edge-effect problems are addressed by augmenting the data within a window with forecasts. In particular, as in Dagum (1987), Stock and Watson (1999), Kaiser and Maravall (1999), Mise, Kim and Newbold (2005), and Clark and Kozicki (2005) – and as is done routinely in seasonal-adjustment procedures – one should augment the window sample data with forecasted data. In the situation depicted in Figure A.1 we have κ sample data points (from time period s to time period s+κ), and m months of projections, yielding a (κ+m)-quarter window (from time period s to time period s+κ+m). We then use a two-sided filter to partition that window into persistence components, and then save the partition at date s+κ; notice that this is a one-sided partition, since no *data* after date s+κ are used. To obtain the partition at date s+κ+1,

we repeat this procedure, obtaining a forecast from data s+κ+1 to data s+κ+1+m, then use a two-sided filter over dates s+1 to s+κ+m+1 and saving the partition at date s+κ+1. This procedure also gracefully allows one to use real-time data.

A2.2 Testing for persistence-dependence

How does testing proceed? In the present case, we wish to test whether the Phillips curve is persistence-dependent. Thus, we partition the unemployment rate *un* into three components (say): un^1, un^2, and un^3. Then we replace *un* in the Phillips curve specification with its 3 components. One may readily test for persistence-dependence using a standard Chow test. Since the components sum to the original series and are based upon one-sided filtering, the causality structure and the properties of the error term are preserved. For more details, see the appendix to Ashley, Tsang, and Verbrugge (2020).

A2.3 Sensitivity to forecasts and filter

Ashley and Verbrugge (2022b) demonstrate that the resultant persistence decomposition is not very sensitive to the number of forecast periods chosen, as long as at least a year of projections are used, nor to the frequency filter used (the Iacobucci-Noullez filter, the Christiano-Fitzgerald filter, or the Ashley-Verbrugge filter) nor to the details of how these forecasts are produced (as long as they are reasonably accurate).

What is crucial is to *partition* the explanatory variables into an interpretably small set of frequency/persistence components that add up to the original data, using moving windows passing through the data so that the filtering is done in a backward-looking or one-sided manner. The Ashley-Verbrugge filter has a key advantage: it can partition the time series into *k* components in a single pass and is thus more readily used for discovering the persistence-dependence in the original data. Other filters must be used in an iterative manner, and in our experience, results are disappointing if one attempts to partition the data into more than 3 components. Furthermore, Ashley and Verbrugge (2022b) demonstrate that the results using other filters are somewhat sensitive to the manner in which this iteration is done.

But with these details in mind, what is of practical macroeconometric importance is to allow for frequency/persistence dependence in the relationship, not – so long as one is mindful of the basic desiderata delineated above – the technical details of precisely how the explanatory variable is partitioned into its frequency/persistence components. Ashley and Verbrugge (2023) report that alternative techniques usually yield quite similar empirical results in practice; see Ashley and Verbrugge (2022b) for more details. RATS, Stata, and Matlab code to accomplish this type of one-sided decomposition (using simple univariate or multivariate forecasts) is available from the authors.

A2.4 Rationale for partitioning, one-sided filtering, and filtering only explanatory variables

Why are partitioning, one-sided filtering, and restriction of the filtering solely to the $u_t - u_t^*$ data all essential? Partitioning is necessary to ensure that these 3 components of the unemployment rate gap add up to the original data, making it easy to test the null hypothesis that the coefficients with which these 3 components enter a regression model for the inflation rate are all equal; it also allows for straightforward interpretation. One-sided filtering is necessary because two-sided filtering – such as ordinary HP filtering or ordinary spectral analysis – inherently mixes up future and past values of the unemployment rate gap in obtaining the persistence components, distorting the causal meaning of inference in the resulting inflation model and limiting its use for practical forecasting and/or policy analysis. These distortions from the use of two-sided filtering are particularly severe when the dependent variable is also filtered and when the key relationship likely (as here) involves feedback from the dependent variable (inflation) to the (filtered) components of $u_t - u_t^*$ being used as explanatory variables. Fundamentally, this is because filtering the dependent variable in a regression model implies that the model error term is similarly filtered. For more details, see Ashley and Verbrugge (2008, 2022a); for a "practical" comparison of methods, including the Hamilton (2018) filter, see Ashley and Verbrugge (2022b).

How about two-sided spectral estimates or filtering with wavelets? These are two-sided methods, so the same criticisms apply. Hence, two-sided cross-spectral estimates or filtering with wavelets are ruled out for analyses of the present sort. And regarding spectral methods, even absent feedback, transfer function gain and phase plots are substantially more challenging to interpret than our approach; even without the presence of feedback, Granger describes interpretation of such plots as "difficult or impossible" (Granger, 1969).

Appendix B: Identification

We adopt the Swanson and Granger (1997) (SG) approach to identification.[2] This method is built upon the fact that most structural causal models, whether linear or nonlinear, imply overidentifying constraints. In particular, a given structural model implies correlation and partial correlation constraints on reduced-form regression residuals ($e_{X,t}, e_{Y,t}, e_{Z,t}$). A partial correlation is the conditional correlation between two variables, conditioned on one or more other variables. For instance, the partial correlation of $e_{Y,t}$ and $e_{Z,t}$ after removing the influence of $e_{X,t}$ is

$$\rho\left(e_{Y,t}, e_{Z,t} \mid e_{Xt}\right) = \frac{\rho\left(e_{Y,t}, e_{Z,t}\right) - \rho\left(e_{X,t}, e_{Z,t}\right)\rho\left(e_{X,t}, e_{Y,t}\right)}{\left[1-\rho^2\left(e_{Y,t}, e_{Z,t}\right)\right]^{1/2}\left[1-\rho^2\left(e_{X,t}, e_{Y,t}\right)\right]^{1/2}}.$$

Under fairly weak assumptions, such overidentifying constraints may be tested using standard *t*-statistics (see below).

Now suppose one such constraint is rejected. Not only is the model under consideration rejected, but also *all* structural models that share this constraint are rejected. Hence, such tests may be used to *restrict* the class of models that are consistent with the data. By virtue of ruling out candidate models that are inconsistent with the data, tests of such overidentifying constraints thus prove useful in *specifying* a structural model. This procedure substantially reduces the subjective nature in the typical SVAR methodology.

We demonstrate how this works using a simple example. Consider the following structural model, a (linear) SVAR involving 3 variables, *X*, *Y*, and *Z*; for simplicity, assume that each is standardized to have mean 0 and standard deviation 1. The particular model is a structural vector autoregression of order 2:

$$\begin{bmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & 0 & 1 \end{bmatrix}\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & b_{13} \\ 0 & b_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & c_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_{t-2} \\ y_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} v_{X,t} \\ v_{Y,t} \\ v_{Z,t} \end{bmatrix} \quad (1)$$

In matrix notation, the SVAR is denoted

$$A\mathrm{M}_t = B\left(L\right)M_t + V_t$$

where $\mathrm{M}_t \equiv \left(X_t, Y_t, Z_t\right)'$, $B\left(L\right)$ is a matrix lag polynomial, and $V_t \equiv \left(v_{X,t}, v_{Y,t}, v_{Z,t}\right)'$. The corresponding reduced-form model is given by

$$\mathrm{M}_t = A^{-1}B\left(L\right)M_t + A^{-1}V_t \equiv \Phi\left(L\right)M_t + E_t$$

where $\Phi\left(L\right)$ is a matrix lag polynomial, and $E_t \equiv \left(e_{X,t}, e_{Y,t}, e_{Z,t}\right)'$. Identification of the SVAR implies obtaining estimates of *A* and *B*(*L*), along with the diagonal variance-covariance matrix *V*. We assume that structural residuals are distributed normally.

[2] This method builds upon work in causal modeling (e.g., Glymour and Spirtes, 1988) and is extended in Demiralp and Hoover (2003) and Demiralp, Hoover and Perez (2008); see also Moneta (2008). The method originated in Blalock (1961).

We now *represent* this model using a causal graph, in Figure A.2. In such figures, an arrow denotes a causal influence: a solid arrow represents a within-period influence, while a dashed arrow represents an intertemporal influence. Notice that time *t* variables are depicted as a function of other time *t* variables and lagged variables. For simplicity, the influence of the exogenous and independent structural shocks v_t^k on variables $k \in \{X, Y, Z\}$, are not depicted. This causal graph is a directed acyclic graph (DAG): it is directed because arrows lead from one variable to another variable, and it is acyclic because one cannot start from any variable, following arrows, and return to that starting variable. Causal graphs are useful tools since they allow one to quickly deduce whether a model can be identified (if all variables are observed, it suffices that it is directed and acyclic), and how the model can be tested. In particular, it is very easy to *read* the implied partial correlation overidentifying restrictions for a given structural model, when the model is expressed as a graph. We do this momentarily.

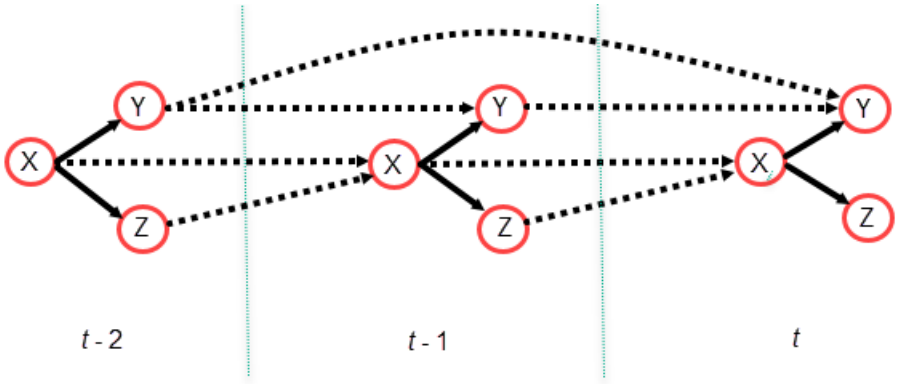


Figure A.2: A structural VAR, with causal influences depicted. Solid lines depict contemporaneous causation; dashed lines depict intertemporal causation. Thus, for example, at time *t*, variable *Y* is influenced by variable *X* contemporaneously, by its own value at time *t*-1, and by its own value at time *t*-2.

VAR modeling begins by estimating the reduced-form model, yielding the residuals (e_X, e_Y, e_Z). How may tests conducted upon these residuals be used to reject models? If Figure A.1 depicts the data-generating process, then we can immediately deduce many partial correlation restrictions amongst those residuals. For instance, this model implies that $\rho\left(e_{X,t}, e_{Y,t}\right) \neq 0$; $\rho\left(e_{X,t}, e_{Z,t}\right) \neq 0$; $\rho\left(e_{Y,t}, e_{Z,t}\right) \neq 0$; and $\rho\left(e_{Y,t}, e_{Z,t} \mid e_{X,t}\right) = 0$. The associated test statistics are standard t-statistics from a least-squares regression of this sort: $e_{Y,t} = \alpha_1 e_{Z,t} + \alpha_2 e_{X,t} + \varepsilon_t$. In the latter regression, the model in Figure A1 would be rejected if $\hat{\alpha}_1 \neq 0$.

This procedure allows us to test a *given* model; but how do such partial correlations help us to deduce the structural model to begin with? Let us return to the SVAR context above, where

identification consists of placing restrictions on the A matrix in equation (1).[3] With three variables, there are 22 possible models (6 of which correspond to Cholesky identification schemes – see Figure F.1 in Appendix F). A researcher not making use of the SG approach would have to use auxiliary information, such as economic theory and timing restrictions, to select one of these 22 models. Now suppose that the DGP is Equation (1). As we now informally demonstrate (the procedure is described somewhat more formally below), the SG approach would allow us rule out 19 of the models, leaving only 3 models to choose between. Clearly, this greatly reduces (though does not fully eliminate) the role of subjective elements in identification.

In the first step, we test all pairwise correlations among the regression residuals. This would indicate that the residuals are all pairwise correlated. This rules out the last 7 models in Figure F.1, namely those in which at least one variable is neither caused by, nor causes, any other variable contemporaneously. As the next step, one would conduct all pairwise conditional residual tests, i.e., one would test $\rho\left(e_{Y,t}, e_{Z,t} \mid e_{X,t}\right) = 0$, $\rho\left(e_{X,t}, e_{Z,t} \mid e_{Y,t}\right) = 0$, and $\rho\left(e_{Y,t}, e_{X,t} \mid e_{Z,t}\right) = 0$ using OLS regressions. Given the data generating process, the last two hypotheses would be rejected, but not the first. This rules out 12 of the remaining models, namely those in which Y and Z have a relationship that are not intermediated in some fashion by X. Putting this differently, these tests would tell us that there are *only three* possible models that are compatible with the data: those in which $Y \rightarrow X \rightarrow Z$, or $Z \rightarrow X \rightarrow Y$, or $Y \leftarrow X \rightarrow Z$. Only at this point would one use prior economic information (in the usual manner) to select from amongst the three candidate models. But these residual tests are useful here as well; for instance, economic theory can sometimes pin down the direction of causality based upon the *sign* of the partial correlation. Alternatively one can use the usual timing restrictions – bearing in mind that at the micro level, agents may be responding to the micro data they currently observe, data that will later be aggregated up to data published by a statistical agency.

The heuristic search procedure involves three steps and relies upon the weak "faithfulness" assumption that if X causes Y (or vice versa) within the period, then their residuals will be correlated.[4] First, compute all bivariate partial correlations, and examine their statistical significance. If the correlation between e_X and e_Y is weak, and $\rho\left(e_{Y,t}, e_{X,t}\right) = 0$ cannot be

[3] Of course, each unique set of restrictions on the A matrix corresponds to an entire class of models, wherein intertemporal relationships are not restricted. However, intertemporal relationships may be estimated without ambiguity from the data, so identification consists of restrictions on the A matrix. For brevity, we refer to the class of models corresponding to a particular structure of the A matrix as "a" model.

[4] We depart somewhat from Swanson and Granger (1997), who forego unconditional correlation tests and instead only examine all conditional correlations; this evidently mitigates reliance on the faithfulness assumption. The faithfulness assumption will fail under "measure-zero" cases where X causes Y, but the two variables are uncorrelated because X causes Z and Z causes Y, and the two causal paths exactly cancel. In the literature, the "faithfulness-failure" examples occur when there is a *decisionmaker* who specifically exerts control over variable Z to accomplish this "cancelation." If there is reason to believe that such a situation exists in a given context, the decisionmaker should be included in the causal graph, and the unconditional correlation test would be interpreted through this lens. In the general case, we recommend the usage of unconditional correlation tests, following standard practice in the causal analysis literature (see, e.g., Moneta 2008). Note that in practice, what matters most for impulse response function estimates are the identifying assumptions made vis-à-vis variables whose residuals are strongly correlated.

rejected, then the data reject $X \rightarrow Y$ and $Y \rightarrow X$ within the period. In an SVAR, the corresponding entries in the impact matrix A would be set to 0.

Second, for those variable pairs (X, Y) with significant correlation, construct trivariate partial correlations with all third variables Z, paying particular attention to those that are correlated with both. If $\rho\left(e_{Y,t}, e_{X,t}\right) = 0$ can be rejected, but if $\rho\left(e_{Y,t}, e_{X,t} \mid e_{Z,t}\right) = 0$ cannot be rejected, then we again conclude that the data reject $X \rightarrow Y$ and $Y \rightarrow X$; their correlation stems from a joint relationship with Z. Third, construct all models that are consistent with this evidence, and select the one that is in accord with economic theory priors.[5] In our experience (and in the experience of Granger and Swanson), parsimonious models appear to agree with the data in most cases, and economic theory often plays a minor role in the selection of the final model.[6] (In models with numerous variables, one may formally test higher-order partial correlation constraints implied by the model.)

While a joint testing procedure is unavailable, so that the usual size problems might arise, in practice this issue is often moot. This is because in many cases, the significance levels of tests can be adjusted significantly without any change in inferences. Furthermore, when a borderline case is "accommodated" – i.e., if the model that, owing to a modest partial correlation, initially ruled out a contemporaneous causation between X and Y, is then extended to specify either $X \rightarrow Y$ or $Y \rightarrow X$, estimation typically yields impulse response functions that are insensitive to this choice.

After a model has been selected, one can perform a final test of joint significance of correlations by estimating the A matrix in the usual manner via maximum likelihood. This test might indicate that a particular entry, a_{ij}, is not statistically different from 0. At that point, one might decide to exclude it.

[5] In models with many variables, there are higher-order tests that may be useful.
[6] If more than one model appears equally reasonable, one may investigate the sensitivity of, e.g., IRFs, to model choice.

Appendix C: Model Estimation Results

$$\pi_t^{PPI} = \alpha^{PPI} + \sum_{j=1}^{4} \beta_j^{PPI} \pi_{t-j}^{PPI} + \delta \Delta u_{t-1}^{medfreq} + e_t^{PPI} \tag{1}$$

$$\begin{aligned} \pi_t^{CoreG} = {} & \alpha^{CoreG} + \phi_1^{CoreG} \pi_{t-1}^{CoreG} + \phi_2^{CoreG} \pi_{t-4}^{CoreG} + \phi_5^{CoreG} \pi_{t-5}^{CoreG} + \\ & + \beta_0^{CoreG} \pi_t^{PPI} + \beta_1^{CoreG} \pi_{t-3}^{PPI} + \lambda^{CoreG} u_{t-4}^{+medfreq} + \psi I^{1995} + v_t^{CoreG} \end{aligned} \tag{7}$$

Table A.1: Regression Results for Equations (1) and (7).

Regressor	Eq. (1); Dependent variable π_t^{PPI} Coefficient estimate	Standard error	Eq. (7); Dependent variable π_t^{CoreG} Coefficient estimate	Standard error
α	0.26	0.13	-0.13	0.05
π_{t-1}^{CoreG}			1.00	0.04
π_{t-4}^{CoreG}			-0.60	0.08
π_{t-5}^{CoreG}			0.46	0.07
π_t^{PPI}			-0.00	0.01
π_{t-1}^{PPI}	1.49	0.08		
π_{t-2}^{PPI}	-0.78	0.15		
π_{t-3}^{PPI}	0.04	0.15	0.03	0.01
π_{t-4}^{PPI}	0.11	0.08		
$u_{t-4}^{+medfreq}$			-0.26	0.22
$\Delta u_{t-1}^{medfreq}$	-5.85	1.57		
I^{1995}			0.35	0.13
$\overline{R}^2$	0.87		0.96	

Note: We elected to retain $u_{t-4}^{+medfreq}$ and π_{t-5}^{CoreG} in the core goods equation since, absent I^{1995}, both clearly belong in the regression. Dropping π_{t-5}^{CoreG} would have little influence, given the size of the estimated coefficient, but dropping $u_{t-4}^{+medfreq}$ would make inflation a tad less responsive to recessionary pressure.

Table A.2: Regression Results for Equations (4) and (8).

$$\pi_t^{MServXH} = \alpha^{MServXH} + \gamma_1^{MServXH}\pi_{t-1}^{MServXH} + \gamma_2^{MServXH}\pi_{t-2}^{MServXH} + \gamma_5^{MServXH}\pi_{t-5}^{MServXH} + \\ + \beta_0^{MServXH}\pi_t^{PPI} + \lambda^{MServXH}u_{t-1}^{+medfreq} + \mu^{MServXH}u_{t-1}^{-lowgap} + v_t^{MNHServ} \quad (8)$$

$$\pi_t^{Hous} = \alpha^{Hous} + \sum_{j=1}^{5}\eta_j^{Hous}\pi_{t-j}^{Hous} + \beta_0\pi_t^{MServXH} + \lambda^{Hous}u_{t-1}^{+medfreq} + \mu^{Hous}u_{t-4}^{-lowgap} + e_t^{Hous} \quad (4)$$

Regressor	Eq. (8); Dependent variable $\pi_t^{MNHServ}$		Eq. (4); Dependent variable π_t^{Hous}	
	Coefficient estimate	Standard error	Coefficient estimate	Standard error
α	0.02	0.03	0.24	0.06
$\pi_t^{MServXH}$			0.01	0.02
$\pi_{t-1}^{MServXH}$	1.19	0.11		
$\pi_{t-2}^{MServXH}$	-0.35	0.12		
$\pi_{t-5}^{MServXH}$	0.15	0.06		
π_{t-1}^{Hous}			1.25	0.08
π_{t-2}^{Hous}			-0.26	0.13
π_{t-3}^{Hous}			0.06	0.13
π_{t-4}^{Hous}			-0.49	0.13
π_{t-5}^{Hous}			0.36	0.07
π_t^{PPI}	0.00	0.00		
$u_{t-1}^{-lowgap}$	-0.13	0.04		
$u_{t-4}^{-lowgap}$			-0.15	0.06
$u_{t-1}^{+medfreq}$	-0.19	0.08	-0.88	0.15
$\overline{R}^2$	0.98		0.97	

Table A.3: Regression Results for Equations (5) and (9).

$$u_t^{medfreq} = \sum_{j=1}^{2} \lambda_j^{med} u_{t-j}^{medfreq} + \sum_{j=1}^{4} \mu_j^{med} u_{t-j}^{lowgap} + \beta^{med} \pi_{t-1}^{PPI} + e_t^{medfreq} \quad (5)$$

$$u_t^{lowgap} = \alpha^{lowgap} + \sum_{j=1}^{2} \mu_j^{low} u_{t-j}^{lowgap} + \sum_{j=0}^{4} \lambda_j^{low} u_{t-j}^{medfreq} + \sum_{j=1}^{4} \beta_j^{low} \pi_{t-j}^{PPI} + v_t^{lowgap} \quad (9)$$

Regressor	Eq. (5); Dependent variable $u_t^{medfreq}$		Eq. (9); Dependent variable u_t^{lowgap}	
	Coefficient estimate	Standard error	Coefficient estimate	Standard error
α	0.00	0.00	0.01	0.01
π_{t-1}^{PPI}	0.00	0.00	-0.01	0.01
π_{t-2}^{PPI}			0.02	0.01
π_{t-3}^{PPI}			-0.01	0.01
π_{t-4}^{PPI}			0.01	0.01
u_{t-1}^{lowgap}	0.49	0.03	0.58	0.14
u_{t-2}^{lowgap}	-0.47	0.04	0.37	0.14
u_{t-3}^{lowgap}	-0.24	0.05		
u_{t-4}^{lowgap}	0.23	0.04		
$u_t^{medfreq}$			1.65	0.22
$u_{t-1}^{medfreq}$	0.91	0.06	-1.04	0.26
$u_{t-2}^{medfreq}$	-0.28	0.06	1.28	0.26
$u_{t-3}^{medfreq}$			-0.72	0.24
$u_{t-4}^{medfreq}$			0.43	0.14
$\overline{R}^2$	0.97		0.99	

Appendix D: Historical Forecasts

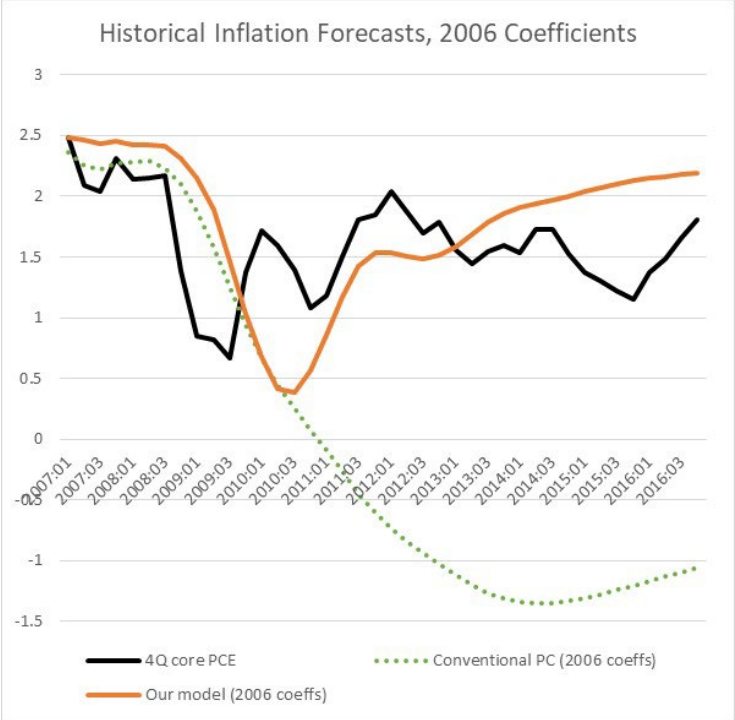
To highlight the fit of our model to historical core PCE inflation data, in Figure D.1 below, we plot an inflation forecast from our model, conditional on the actual path of the unemployment rate over the 2007Q1-2016Q4 period, alongside the conditional forecast from a more conventional Phillips curve model. This latter model is specified as a *linear* bivariate model of core PCE inflation and the unemployment gap.[7] For forecasts displayed in Figure D.1(a), both models are estimated using data through 2006Q4, and coefficients' estimates are assumed fixed through the forecast sample; inflation forecasts are constructed recursively from 2007Q1 through 2016Q4, so neither model conditions on post-2006 inflation data. For forecasts displayed in Figure D.1(b), both models are estimated using data through 2019Q4; but inflation forecasts are still constructed recursively from 2007Q1 through 2016Q4, so neither model conditions on post-2006 inflation data.

Figure D.1(a) corresponds more closely to a real-time forecast exercise, since the ten-year forecasts do not condition on the post-2006 evolution of inflation. The forecast from our model displays a sharp inflation deceleration followed by a strong partial rebound starting in 2011, broadly capturing the dynamics of core PCE inflation over the period.[8] In contrast, the conventional model features the well-known "missing disinflation" puzzle. The pre-2006 linear Phillips curve relationship is estimated to be strong enough such that the large unemployment gap during the Great Recession implies notable downward force on inflation forecast. Figure D.1(b) displays "forecasts" which use post-2006 inflation data to estimate coefficients, but still generate inflation forecasts recursively, i.e., without reference to the evolution of actual inflation after 2006. Our model's forecasts using circa-2019 coefficients is broadly similar to the ones constructed using circa-2006 coefficients – evidently, the Phillips curve coefficients are only slightly weakened – and differences mainly relate to long-run convergence levels of the components. Conversely, the linear Phillips curve estimated in 2019 solves the missing disinflation puzzle by weakening the Phillips curve and thereby introducing the "downward speed" puzzle, namely this model almost entirely misses the deceleration of inflation early in the forecast. We think this figure convincingly illustrates the superior fit of our nonlinear Phillips curve model specification to inflation data compared to the linear analogue

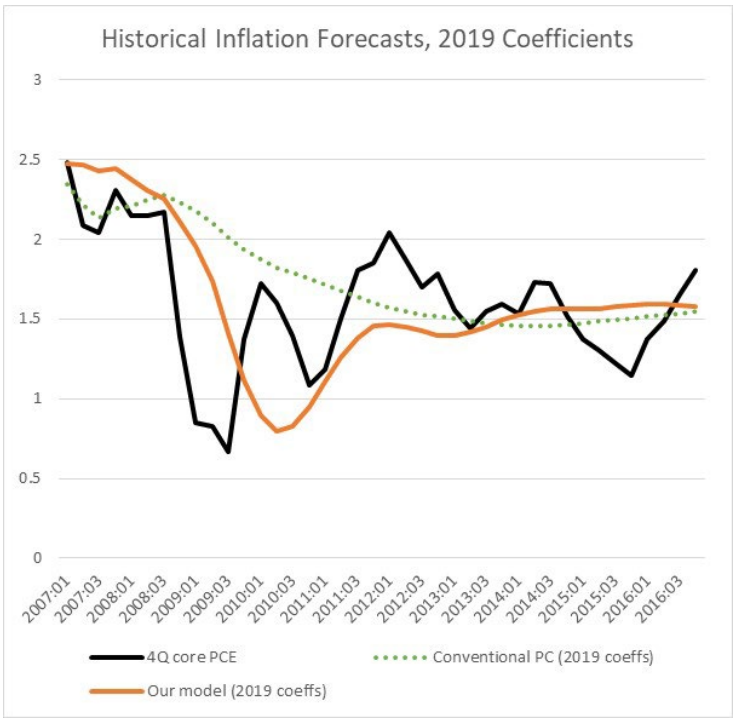
As discussed in Clark and Zaman (2013), these sorts of conditional forecasting exercises provide us some indication of how well a model is formulated. By conditioning on the historical path of the unemployment rate (thereby removing this source of uncertainty), we can assess how well the rest of the model (i.e., the inflation subcomponents) translates this information into inflation pressures and correctly captures the persistence of inflation. If the model is well-constructed and the unexpected shocks to the inflation components are not too large over the forecast horizon, then the conditional forecast of inflation should be close to the actual inflation path. As noted above, the conditional forecast of inflation coming from our nonlinear model does a respectable job tracking actual inflation over a 10-year period, and clearly outperforms its more standard

[7] The regressors of the model are a constant, five lags of 4Q core PCE, and the unemployment gap.
[8] The forecast misses from 2008-2010 are largely attributable to core PCE being driven by large outliers in non-market services, as discussed in Verbrugge (2022).

counterpart, lending strong support to our model and providing reassurance about its ability to accurately provide conditional forecasts of the sort conducted in this study.



(a) Using coefficients estimated in 2006Q4



(b) Using coefficients estimated in 2019Q4

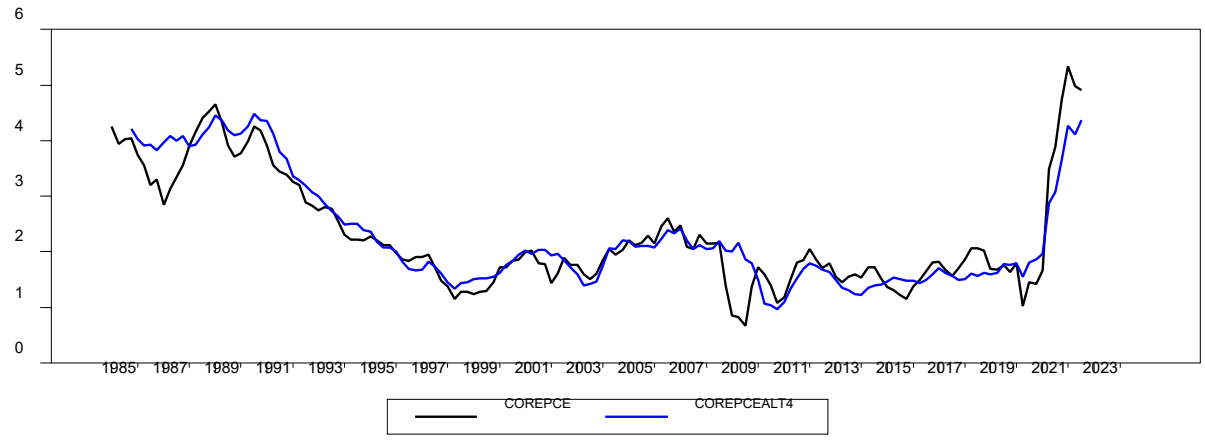
Figure D.1: Historical 10-year forecast from the model and from a conventional Phillips curve. Both are conditional recursive forecasts. The models see no inflation data after 2006Q4, but the forecasts are conditioned on the evolution of the unemployment rate.

Appendix E: An Alternative to Core PCE

Implicitly, our work defines an alternative to core PCE, namely the inflation series defined by

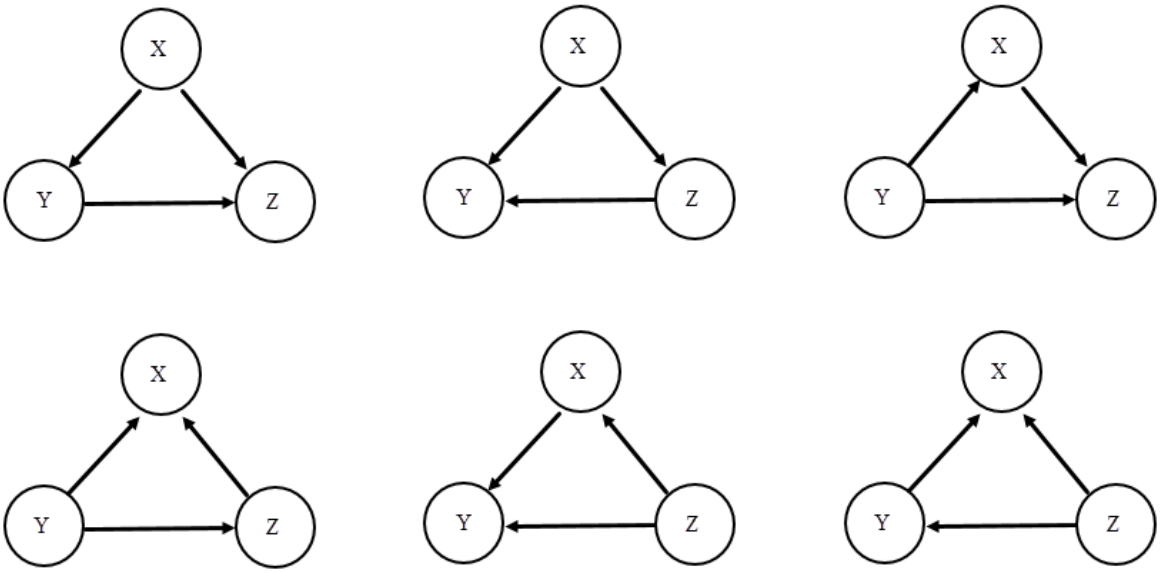
$$\pi_t^{CorePCEAlt} = w_t^{CoreG}\pi_t^{CoreG} + w_t^{Hous}\pi_t^{Hous} + w_t^{CoreServExHous}\pi_t^{MedianCoreServExHous}$$

The 4Q version of this series is plotted below, alongside its core PCE counterpart.

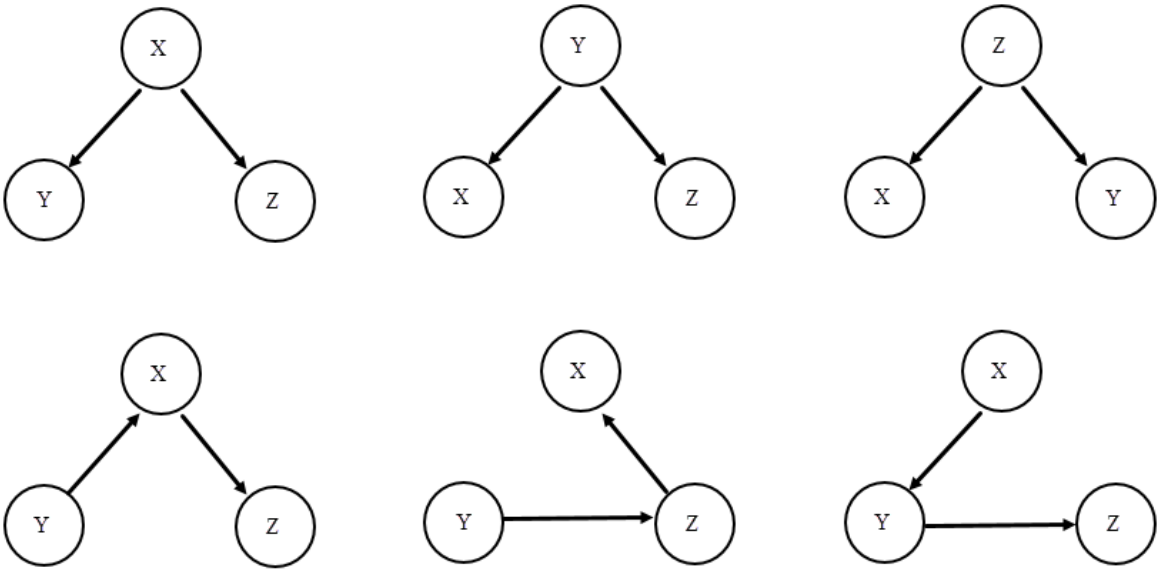


The alternative series has a number of attractive features as an alternative to core PCE. First, it is evidently smoother than core PCE. Second, it lacks the problematic dynamics of core PCE during the Financial Collapse, dynamics that are highlighted by Verbrugge (2022). Third, it has a close connection to core PCE (e.g., it is an approximately unbiased relative to core PCE), so it would be much easier for devotees of core PCE to switch to. Finally, it lends itself to most of the storytelling that is a perceived advantage of core PCE over, say, trimmed-mean PCE.

Appendix F: Enumeration of all identifiable models for 3-variable SVAR



Cholesky identification models



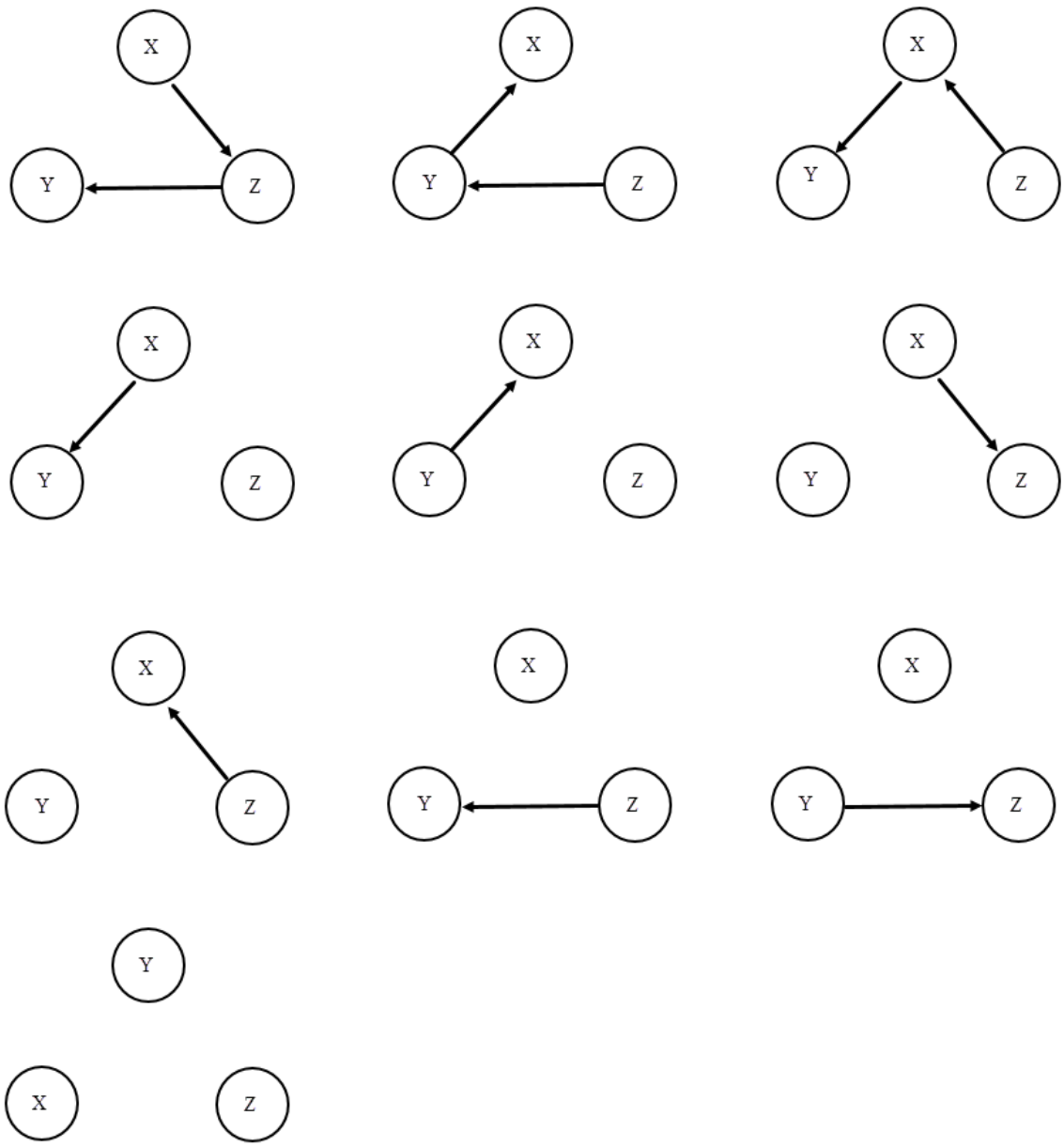


Figure F.1: All Identifiable Structural Models Involving Three Variables *X*, *Y*, and *Z*.

Appendix G: PPI-IG as an Indicator of Supply Shocks

In this Appendix section, we present information suggesting that movements in PPI-IG are dominated by supply shocks, even during the recent period, although demand shocks also play a non-negligible role as well.

Clearly, PPI-IG is an equilibrium price, and so (in general) its movements will reflect the influence of both demand and supply shocks. One must make identifying assumptions to isolate the effect of supply on the PPI-IG.

Theoretically, under our identifying assumptions, the PPI-IG, in conjunction with the other variables in the model, allows us to identify supply shocks – as the structural residuals in the model. This relies on our assumption that activity, as measured by the unemployment rate, is an adequate proxy for demand influences on the PPI-IG. But we now explore the robustness of this assumption.

The New York Fed's approach to removing demand-side effects from their Global Supply Chain Pressures Indicator involves regressions on the PMI New Orders index (PMI-NO).[9] Note that this variable is a very influential predictor of industrial output over the next several months. Thus, we first investigated the extent to which, in *monthly* data, the PMI-NO index influenced the PPI-IG. If PMI-NO are an important driver of contemporaneous PPI-IG movements, we might be more reluctant to assert that the structural shocks hitting the PPI-IG are predominately supply shocks.

Our first test was to regress 12-month PPI-IG on *contemporaneous* PMI-NO, as well as 4 additional lags of this variable (and we find lags 1, 2, and 12 and 13 seem appropriate in this regression), and one lag each of the change in the medium-frequency component, and the change in the low-frequency component, of unemployment. We do find a statistically-significant contemporaneous impact of PMI-NO, and it is positive (suggesting that PMI New Orders cause PPI-IG, consistent with a demand influence on PPI-IG). However, the estimated coefficient is small, so its economic magnitude is rather modest: a one-standard deviation in PMI -NO leads PPI to rise contemporaneously by 0.17 ppts – but this series has a standard deviation that exceeds 3.

Our second test, reported in Verbrugge and Zaman (2023a), was a rather involved robustness text. The model in that paper included trimmed mean PCE, rather than the components of core PCE as in the present paper, but was otherwise identical. In this test, we respecified the baseline four-variable model to include PMI-NO as an additional variable. This required a new specification search for each equation, followed by a new application of our identification apparatus, etc. We sought to answer four chief questions: a) were there notable changes in the historical forecast performance of the model over the recent period?; b), were there notable changes in the impulse response function of inflation to a PPI-IG shock?; c), what is the

[9] FRBNY also uses a quantities index to adjust some of the underlying series, but the only measure of the quantities index we have access to is the Markit measure, and this series starts in 2009. Note that PMI New Orders is not, strictly speaking, a "demand shock," but rather an equilibrium variable subject to both demand and supply influences. Thus, regression on PMI-New Orders removes more than simply demand shocks.

responsiveness of PPI-IG to a shock to PMI-NO?; and d), historically, how important are PMI-NO shocks to the variance of PPI-IG?

We found that lags of PMI-NO influenced PPI-IG and both unemployment components, and that PMI-NO responded to lags of PPI-IG and the low frequency gap component. After completing our specification search, we identified the model. This revealed contemporaneous relationships between PMI-NO and PPI-IG, and also between PMI-NO and the low frequency gap. Thus, we had to make a causality assumption for PMI-NO vis-à-vis PPI-IG, and also for PMI-NO vis-à-vis the low-frequency gap. To be conservative (i.e., to give "demand" as big a role as possible), we assumed that PMI-NO caused both. This implies that there is no role in this new model for unemployment to influence PMI-NO contemporaneously, removing one source of "demand" as a driver of PMI-NO shocks. One should bear in mind that a PMI-NO shock is presumably a mixture of both demand and supply influences.

We then investigated question a). Our specification and identification procedures indicated that there was no direct influence of PMI-NO on trimmed-mean PCE inflation. Thus, the model's historical forecast performance was unaffected by the addition of PMI-NO.

As to question b), we discovered that the IRF of trimmed mean PCE inflation to a PPI-IG shock was moderated, but only very slightly. To the extent that both unemployment and PMI-NO capture the influence of demand on PPI-IG, this implies that this IRF depicts the response of trimmed mean PCE inflation to a supply shock – and our original IRF is pretty close to this. This was very reassuring to us. Qualitatively, all of the conclusions in the paper about this IRF remain unchanged.

Regarding question c), we found that PMI-NO does indeed influence PPI-IG in quarterly data. But how important is this linkage? First, we look at the response of PPI-IG to shocks. A one-standard deviation shock in PMI-NO eventually causes PPI-IG to rise, with a peak response about one year after the shock. The rise is about one half of a standard deviation (of PPI-IG). Conversely, a one standard deviation shock in PPI-IG itself eventually causes PPI-IG to rise by 1.4 standard deviations. (Also, neither unemployment rate component is a quantitatively important driver of PPI-IG.) Over our estimation period, the correlation of the residuals in the PPI-IG equation with versus without PMI-NO is 0.94; only rarely is there a big gap between the two residual series. This suggests that PMI-NO is not generally an important driver of PPI-IG.

Regarding question d), in our augmented model that includes PMI-NO, if we shut down PMI-NO shocks (over our estimation period), the variance of PPI falls by about 20%. This suggests that PMI-NO is a statistically significant driver of PPI-IG, but that most of the variance in PPI-IG is driven by supply shocks.

Finally, we investigated whether the recent period was different. PMI New Orders did rise sharply in the COVID recovery, but they essentially remained within the range experienced over our estimation sample period (1985-2019) – suggesting that the main driver of the very large PPI shocks over this period seems to have been supply. (It is also the case that during this period, the medium-frequency component of the unemployment rate attained historic lows, then rose rapidly. Given its influence on PPI-IG, the influence of rapidly-changing demand on PPI-IG was arguably captured, at least to some extent.) However, our augmented model allows us to take a

deeper dive. Using counterfactual simulations, we turned off PMI-NO shocks from 2020:1 onwards, to explore the impact of these shocks on PPI-IG. This exercise suggested, perhaps not surprisingly, that PMI-NO shocks were rather important drags on PPI-IG in 2020. (This is consistent with the findings of Meyer, Prescott and Sheng, 2022.) After that, they provided upward pressure: in 2021, they boosted PPI-IG by roughly 20%, and in early 2022, the boost rose to roughly 30%. Particularly in the recent period, then, it is important to keep in mind that the PPI-IG is an equilibrium variable; in any given episode, might reflect demand shocks more than in other episodes.

Overall, in these quarterly data, we conclude that PMI-NO shocks are an "important" driver of PPI-IG movements, generally far from the predominant driver, but occasionally quite important.

We view all of this as evidence that, while our baseline model does not completely remove the impact of demand on the PPI-IG structural residual, the majority of the variance in this structural residual "on average" is due to supply shocks. Strictly speaking, however, the PPI-IG structural residuals are not 100% pure indicators of supply shocks.

Appendix H: Nonlinear Impulse Response Functions and Error Bands

For computing the nonlinear IRFs to variable X, we make use of the notion that an IRF is the difference between a forecast with a particular shock, and the forecast without one. In particular, given the model's nonlinear nature, we construct impulse response functions (IRFs), forecasts, and error bands via counterfactual simulations. To obtain accurate error band estimates, we augment the bootstrap procedure outlined in Kilian and Lütkepohl (2017) in three ways. First, motivated by Phillips and Spencer (2011), we estimate and apply a small-sample bootstrap variance correction. (In our context, this turned out to be rather small.) Second, we estimate the bias-corrections for $B(L)$ as in Kilian (1998), prior to bootstrapping. Finally, we take into account parameter estimation uncertainty via a bootstrap-upon-bootstrap method, along the lines of Potter (2000) and Pérez Forero and Vega (2016). For each such bootstrap, we apply the bias corrections previously estimated.

In brief, our procedure is as follows. In the outer bootstrap loop, for M replications, we simulate data from the model (with shocks bootstrapped from the estimated residuals) and re-estimate the model. Estimates are then bias-corrected using our bias-correction estimates; for simplicity, explosive draws are thrown out. In the inner bootstrap loop, for N replications, we use this newly-estimated (bias-corrected) model, with corrected variance, to estimate an IRF as in Kilian and Lütkepohl (2017) – which we describe immediately. We set M =500 and N =500, so that for each IRF, MxN=250,000 simulations. We plot the median response, as well as the 10th and 90th percentiles from the simulations, as explained below.

In the inner "Kilian/Lütkepohl" loop, we first select the set of initial conditions that will be considered (e.g., for the medium-frequency component of the unemployment rate, one case is $u^{medFreq}(t) >0$). (Obviously, prior to bootstrapping, we must identify all dates satisfying the initial conditions.) For each of N replications, we randomly sample from these dates as starting conditions. Then we produce k-period forecasts, conditional on the (bias-corrected) estimated model, and (small-sample-corrected) bootstrap draws from the estimated regression residuals (for each period, drawing the vector of residuals over the four equations) This produces a "baseline" forecast for this initial condition and this set of bootstrap draws. Then we produce a second forecast, where we replace the initial (period s) bootstrap draw for the variable X by a given shock. In particular, we replace the bootstrap draw at date s by δ_X, where δ_X corresponds to a specific shock for X, whose size is chosen to be of sensible magnitude for X. The difference between the two forecasts, for each draw, is the impulse response function for that draw. This IRF is stored; then we begin the next replication. When all replications have been completed, we compute the median and the 10th and 90th percentiles of all the impulse response functions.

Forecast error bands are computed similarly, except that there is no second "δ_X" draw, and we select the initial forecast date, rather than drawing this randomly.