Appendix to

"The CPI–PCEPI Inflation Differential: Causes and Prospects." 2020.
Wesley Janson, Randal Verbrugge, and Carola Conces Binder.
Federal Reserve Bank of Cleveland, *Economic Commentary*, 2020-06.

A. Techniques and Models

Refined five-year moving average

It is well-known that while a five-year moving average filters out *most* of the fluctuations that revert more quickly than five years, they do this imperfectly; some of those fluctuations "leak through." The result is that an ordinary five-year moving average is not smooth. It is possible to improve upon an ordinary five-year moving average by applying two alternative moving averages in succession: a 16-quarter moving average, and then an 8-quarter moving average. (Technically, by examining these filters in the frequency domain, it is easy to show that these filters are nearly identical in terms of the (low) frequencies that they allow to pass through—namely, all fluctuations that last at least five years—while the (16, 8) filter does a vastly superior job of filtering out all fluctuations that last less than five years.) The end result is, effectively, a smooth five-year moving average.

ARMA(p,q) models

An ARMA(p,q) model of the inflation differential is

$$\left(\pi_t^C - \pi_t^P\right) = \sum_{j=1}^{p} \beta_j \left(\pi_{t-j}^C - \pi_{t-j}^P\right) + \varepsilon_t + \sum_{i=1}^{q} c_i \varepsilon_{t-i}$$

where ε_t is a white-noise error term. (An ARMA model of the inflation or price relative ratio is defined analogously.) The lag length p and the number of moving-average terms q are selected by minimizing an information criterion such as the Akaike Information Criterion (AIC) or the Hannan-Quinn Information Criterion (HQ); this approach is a statistical means of "penalizing" the total number of parameters. Such criteria are often used in the forecasting literature, and enhance the forecasting accuracy of the model. In preliminary analysis, we found that the AIC resulted in frequent and large swings in the ARMA parameter selections, which suggests overfitting. We thus decided to use the HQ criterion, which penalizes extra parameters at a slightly higher rate; this also resulted in more accurate forecasts.

Time-varying parameter models

There are many kinds of time-varying parameter models. What distinguishes these models from other models is that the estimated coefficients are allowed to vary over time—in other words, the estimation procedure explicitly allows the estimated coefficient to be different at different periods. The basic specification is

$$\pi_t^C = a_t + b_t \pi_t^P + u_t$$

and then one must specify various details that govern how quickly the coefficients are allowed to change as time moves forward.

Bayesian vector autoregression

A VAR(p) model can be written as

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + A_c + \varepsilon_t$$

where $Y_t = \left(y_{1,t}, y_{2,t}, \ldots, y_{n,t}\right)$ is a data vector of n random variables, $A_c = \left(c_1, c_2 \ldots, c_n\right)$ is a vector of constants, $A_1, A_2, \ldots, A_n$ are $n \times n$ matrices of VAR coefficients, and $\varepsilon_t \sim N\left(0, \Sigma\right)$. In this equation, p indicates the number of lags.

We estimate the VARs using Bayesian shrinkage (hence the term BVAR). This implies that we use an objective statistical approach to estimation that combines the modeler's prior beliefs with the available data. We achieve this by imposing prior restrictions on the parameter estimates. Specifically, we shrink the coefficients of the VAR toward the univariate random walk model with a drift. Doing this gives us an a priori system consisting of random walk processes. The overall degree of shrinkage is controlled by hyperparameter λ. As $\lambda \to 0$, shrinkage increases and the prior dominates, making data less influential in determining the posterior coefficient estimates (with a $\lambda = 0$ prior equals posterior, so that the data are ignored), whereas as $\lambda \to \infty$, the data dominate the prior and influence the posterior estimates to a greater extent (so that we obtain OLS estimates with $\lambda = \infty$). The value assigned to the hyperparameter λ is 0.15. The BVAR we consider in this paper implements the normal-inverted Wishart prior proposed by Kadiyala and Karlsson (1997) and Sims and Zha (1998), which is basically a version of the Minnesota prior introduced by Litterman (1986).

References

Kadiyala, K. Rao, and Sune Karlsson. 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics*, 12(2): 99–132. https://doi.org/10.1002/(sici)1099-1255(199703)12:2<99::aid-jae429>3.0.co;2-a.

Sims, Christopher, and Tao Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, 39(4): 949–968. https://doi.org/10.2307/2527347.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics*, 4(1): 25–38. https://doi.org/10.1080/07350015.1986.10509491.