Bauer and Swanson’s paper is an important contribution to the literature measuring the dynamic effects of monetary policy shocks on the macroeconomy. It shows what can go wrong using apparently exogenous high-frequency monetary policy surprises as instruments for monetary policy shocks, and shows how to correct the faulty estimates and inference by incorporating appropriate control variables. It is also the first paper to use the expanded set of monetary policy surprises compiled by Swanson and Jayawickrema (2022), another important contribution. The culmination of the authors’ work are the monthly impulse responses shown in Figure 8 of the paper. These are the new benchmark impulse response functions for the effect of Federal Reserve monetary policy shocks on the U.S. macroeconomy.

I. Structural Vector Autoregressions and Local Projections with External Instruments

In his 2008 NBER Summer Institute lectures, Stock (2008) showed how an ‘external’ instrument can be used in a vector autoregression to identify a structural shock and its impulse response function. I begin by summarizing the external instrument framework as this makes it easy to explain Bauer and Swanson’s contributions.

Let $Y_t$ denote a vector of time series variables. In Bauer and Swanson’s paper, $Y_t$ contains four variables: the 2-year Treasury bond rate, the logarithm of the index of industrial production, the logarithm of the CPI and the excess bond premium from Gilchrist and Zakrajsek (2012). The structural VAR is

$$Y_t = \alpha + B(L)Y_{t-1} + u_t \text{ with } u_t = S\varepsilon_t.$$ (1)

The elements of $u_t$ are the VAR’s one-period-ahead forecast errors, $\varepsilon_t$ are the structural shocks, and $S$ is a matrix linking $\varepsilon$ and $u$. $\varepsilon_t$ is assumed to be white noise process with a
diagonal covariance matrix. Solving the model backwards shows how \( Y_t \) depends on current
and lagged values of \( \varepsilon \): 
\[
Y_t = \mu + A(L)S\varepsilon_t
\]
where \( A(L) = [I - B(L)]^{-1} \) and \( \mu = [I - B(1)]^{-1}\alpha \).

Let \( Y_{j,t} \) denote the \( j \)th variable in \( Y_t \) (so \( Y_{1,t} \) is the 2-year Treasury bond rate in this paper),
and decompose \( \varepsilon_t \) as 
\[
\varepsilon_t = (\varepsilon_t^{MP}, \varepsilon_t^{Other})
\]
where \( \varepsilon_t^{MP} \) is the scalar monetary policy shock – the shock of interest in Bauer and Swanson’s exercise –
and \( \varepsilon_t^{Other} \) collects the other structural shocks. Let \( \beta \) denote the first column of \( S \) (so that \( \beta \) is the column of \( S \) corresponding to \( \varepsilon^{MP} \)) and \( \Gamma \) denote the other columns of \( S \). Then
\[
Y_t = \mu + A(L)S\varepsilon_t = \mu + A(L)\beta\varepsilon_t^{MP} + A(L)\Gamma\varepsilon_t^{Other}.
\]
The parameters in \( \mu \) and \( A(L) \) are functions of the VAR parameters \( \alpha \) and \( B(L) \). They are
identified, can be estimated using standard methods, and are unimportant for the insights
provided by Bauer and Swanson. Thus, to focus on the main ideas, assume that \( \mu = 0 \) and \( A(L) = I \), so the model becomes static with
\[
Y_t = \beta\varepsilon_t^{MP} + \Gamma\varepsilon_t^{Other}.
\]
With these simplifications, the econometric problem simplifies to estimating the value of \( \beta \) in
equation (2). This is achieved using a scale normalization and an external instrument.

The scale normalization sets the first element of \( \beta \) to unity, that is
\[
Y_{1,t} = \varepsilon_t^{MP} + \gamma'_1\varepsilon_t^{Other}
\]
where \( \gamma'_1 \) is the first row of \( \Gamma \). This normalization says that the monetary policy shock
is measured in the same units as \( Y_{1,t} \), the 2-year Treasury bond rate. Rearranging (3),
\( \varepsilon_t^{MP} = Y_{1,t} - \gamma'_1\varepsilon_t^{Other} \), and substituting this into (2) yields
\[
Y_{j,t} = \beta_j Y_{1,t} + e_t
\]
where \( e_t = (\gamma_j - \beta_j\gamma'_1)\varepsilon_t^{Other} \) is uncorrelated with \( \varepsilon_t^{MP} \). Conveniently, the parameter
of interest, \( \beta_j \), is the coefficient on \( Y_{1,t} \) in (4). Unfortunately though, \( Y_{1,t} \) is correlated with \( e_t \),
so \( \beta_j \) cannot be consistently using OLS in (4). An instrument is needed.

Thus, consider a variable \( z_t \) that is correlated with \( \varepsilon_t^{MP} \) but uncorrelated with \( \varepsilon_t^{Other} \).
Then \( z_t \) will be correlated with \( Y_{1,t} \) and uncorrelated with \( e_t \), so it is a valid instrument for
estimating \( \beta_j \) in (4). Gertler and Karadi (2015) argued that changes in federal funds rates
futures contracts around FOMC announcements satisfied the conditions required for \( z_t \). Their
argument was plausible – indeed I found it persuasive – and several subsequent papers have used these high-frequency monetary surprises as external instruments to estimate the effect of monetary policy shocks.

The resulting IV estimator is

$$\hat{\beta}_j^{IV}(z) = \frac{T^{-1} \sum z_t Y_{1,t}}{T^{-1} \sum z_t Y_{1,t}} = \beta_j + \frac{T^{-1} \sum z_t e_t}{T^{-1} \sum z_t Y_{1,t}}.$$  (5)

Evidently, if $T^{-1} \sum z_t e_t \approx 0$ and $|T^{-1} \sum z_t Y_{1,t}| \gg 0$, then $\hat{\beta}_j^{IV}(z) \approx \beta_j$.

II. Bauer and Swanson’s critique of Monetary Policy Surprises as Instruments

Bauer and Swanson’s critique goes as follows: first they document (following others that they cite) that the high-frequency monetary surprises used by Gertler, Karadi, and others are correlated with information that was publicly available before FOMC meetings. Let $x_t$ represent this information. Over the sampler period considered in these papers, the sample correlation between $z$ and $x$ is significant, that is

$$|T^{-1} \sum z_t x_t| \gg 0.$$  (6)

Second, $x_t$ is potentially correlated with $\varepsilon_t^{Other}$, and thus potentially correlated with $e_t$ in (4). Projecting $e_t$ onto $x_t$ yields the decomposition

$$e_t = \theta x_t + a_t$$

where $\theta \propto \text{cor}(e_t, x_t)$. Substituting this expression for $e_t$ in (5):

$$\hat{\beta}_j^{IV}(z) = \beta_j + \frac{T^{-1} \sum z_t e_t}{T^{-1} \sum z_t Y_{1,t}} = \beta_j + \frac{\theta T^{-1} \sum z_t x_t + T^{-1} \sum z_t a_t}{T^{-1} \sum z_t Y_{1,t}}.$$  (7)

Thus, the empirical finding that $|T^{-1} \sum z_t x_t| \gg 0$ suggests that $\hat{\beta}_j^{IV}(z)$ is likely to deviate significantly from $\beta_j$ in the sample.

Bauer and Swanson’s solution to the problem also has two parts. First, they add $x_t$ as a control variable in (4). The resulting IV estimator is

$$\hat{\beta}_j^{IV}(z^+) = \beta_j + \frac{T^{-1} \sum z_t^+ e_t}{T^{-1} \sum z_t^+ Y_{1,t}} = \beta_j + \frac{T^{-1} \sum z_t^+ a_t}{T^{-1} \sum z_t^+ Y_{1,t}}.$$  (8)
where $z_t^\perp$ is the residual from projecting $z_t$ onto $x_t$. This eliminates the problematic correlation of the instrument with $e_t$. That is, it eliminates the problematic term in the numerator in (7). But, it also changes the denominator, and Bauer and Swanson show $z_t^\perp$ is only weakly correlated with $Y_{1,t}$. Thus, $\beta_j$ is only weakly identified using $z_t^\perp$ and $\hat{\beta}^{IV}_j(z^\perp)$ is a weak-instrument estimator.

The second part of the solution eliminates the weak instrument problem by adding new data, specifically by augmenting the monetary policy surprises to include surprises around speeches given by the Fed Chair. These are the new data developed in Swanson and Jayawickrema (2022). Bauer and Swanson show that these additional data significantly strengthen the instrument.

Taken together, these two modifications produce credible estimates of the effect of monetary policy shocks on $Y$.

III. How can Monetary Policy “Surprises” be Predictable?

I was puzzled by the empirical results on the predictability of monetary policy surprises documented by Bauer and Swanson and earlier papers. Why is it that changes in federal funds or Eurodollar futures contracts are so predictable? The model presented in this paper, and related arguments in Farmer, Nakamura, and Steinsson (2022) helped me understand that this predictability is a form of sampling error. Let me explain using the model presented in Bauer and Swanson’s paper, but incorporating some of the notation that I used above. Let $i_t$ denote the interest rate (or the Eurodollar futures contract in Bauer and Swanson’s empirical analysis) measured just after a FOMC announcement or Federal Reserve Chair speech. Following the notation I used above, let $x_t$ denote information available just before the announcement and let $z_t$ denote the monetary policy surprise. (In Bauer and Swanson’s model, the monetary policy surprise is denoted $mps_t$ and the pre-announcement information is denoted by $(x_t, H_{t-1})$.) The model’s monetary policy surprise is

$$z_t = i_t - \mathbb{E}(i_t|x_t),$$

so $\mathbb{E}(z_t|x_t) = 0$ and the monetary policy surprise is unpredictable. Through the lens of the model, a non-zero value of $T^{-1} \sum z_t x_t$ is sampling error and a ‘statistically significant’ non-zero value of $T^{-1} \sum z_t x_t$ is unlikely. What then explains the unlikely large value of $T^{-1} \sum z_t x_t$ found in the data?
In the model of Bauer and Swanson, $i_t$ and $x_t$ are related by a parameter $\alpha$, whose value changes stochastically through time. Thus

$$\mathbb{E}(i_t|x_t) = \int \mathbb{E}(i_t|x_t, \alpha_t)p(\alpha_t|x_t)d\alpha_t$$

where $p(\alpha_t|x_t)$ is the conditional pdf of $\alpha_t$. Figure 1 in Bauer and Swanson shows that over their sample period, the Fed’s interest rate policy became increasingly more responsive to the economy, which in the model translates into an unusually long sequence of draws from the right tail of the $p(\alpha_t|x_t)$ distributions. In this sense, investors appeared to be systematically surprised by the Fed, despite using optimal predictors.

I take two lessons from this. First, from (7) whether $\hat{\beta}^{IV}_j(z) \approx \beta_j$ in any given sample depends on the sample covariance between monetary policy surprises and $x_t$, that is $T^{-1} \sum z_t x_t$. Thus, it is useful to control for $x_t$ in samples for which $T^{-1} \sum z_t x_t$ is large, even if $\mathbb{E}(z_t x_t) = 0$.

The second lesson is that $\mathbb{E}(z_t x_t) = 0$ means that $T^{-1} \sum z_t x_t$ is likely to be small in future samples, so controlling for $x_t$ in other sample periods may be unnecessary.

**IV. Dynamics and Bauer and Swanson’s Best Practices**

Through most of my discussion I have neglected the dynamics associated with the VAR coefficients in $B(L)$ in (1). Estimation of these coefficients play a role in the Bauer and Swanson’s ‘Best Practices’ (or ‘Lessons Learned’) at the end of their paper. The authors comment that impulse response “Estimates from SVARs tend to be more precise and less erratic that those based on local projections, but the two are qualitatively similar.” This conclusion is based on unrestricted estimates of the VAR and LP coefficients. Two recent papers, Li, Plagborg-Møller, and Wolf (2022) and Plagborg-Møller and Wolf (2021), provide a more nuanced comparison of LP and VAR estimators. Using empirically relevant designs these papers show that LP estimators tend to be less biased but more variable than VAR estimators, so ranking of the estimators depends on the relative weight placed on bias and variance. Consistent with Bauer and Swanson’s comment, a mean squared error criterion generally favors VAR estimators. That said, shrinkage or Bayes methods improve upon the unrestricted estimators, providing significantly lower mean squared error.
References


