# 4 Structural VAR Tools

This chapter provides an overview of the uses one can put a structural VAR to. We review the construction of structural impulse responses, forecast error variance decompositions, historical decompositions, forecast scenarios, and counterfactual outcomes including policy counterfactuals.

Consider the structural VAR($p$) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t,$$

where the $K \times 1$ vector $y_t$ is presumed to be zero mean for expository purposes. The dimension of $B_i$, $i = 0, \ldots, p$, is $K \times K$. The $K \times 1$ vector $w_t$ is assumed to be white noise. The model is structural in that the elements of $w_t$ are mutually uncorrelated and have clear interpretations in terms of an underlying economic model. We further postulate that the $K$ model variables are driven by $K$ distinct shocks such that their variance-covariance matrix $\Sigma_w$ is of full rank. Thereby we rule out that the structural model includes equations that are merely identities, as is common in traditional simultaneous equations models, rather than being subject to stochastic errors. We also rule out that the data are generated by economic models in which there are fewer than $K$ structural shocks. This assumption excludes, for example, the standard real business cycle model, in which all macroeconomic aggregates are driven by a technology shock only such that the covariance structure of the data is singular.

This model can be expressed in reduced form as

$$y_t = \underbrace{B_0^{-1} B_1}_{A_1} y_{t-1} + \cdots + \underbrace{B_0^{-1} B_p}_{A_p} y_{t-p} + \underbrace{B_0^{-1} w_t}_{u_t}.$$

We normalize the covariance matrix of the structural errors $\mathbb{E}(w_t w_t') \equiv \Sigma_w = I_K$ without loss of generality such that the reduced-form error covariance matrix is $\mathbb{E}(u_t u_t') \equiv \Sigma_u = B_0^{-1} B_0^{-1\prime}$. Going back and forth between the structural and the reduced-form representation requires knowledge of the matrix $B_0$ that governs the instantaneous relationships among the model variables or of its inverse, the structural impact multiplier matrix $B_0^{-1}$. Given that $u_t = B_0^{-1} w_t$, this matrix allows us to express the typically mutually correlated reduced-form innovations ($u_t$) as weighted averages of the mutually uncorrelated structural innovations ($w_t$), with the elements of $B_0^{-1}$ serving as the weights.

Much of the discussion in subsequent chapters of this book addresses the question of how to recover estimates of $B_0^{-1}$ (and, hence, estimates of the structural model coefficients $B_0, \dots, B_p$) from estimates of the reduced-form VAR model, as discussed in Chapters 2, 3, and 5. In the current chapter we focus on the question of how to present the estimates of the structural model in ways that facilitate the interpretation of economic data. For expository purposes, we postulate for now that $B_0^{-1}$ (and hence $B_0$) is known. A review of how to specify and estimate $B_0$ and $B_0^{-1}$ in practice can be found in Chapters 8, 9, 10, 11, 13, 14, and 15.

## 4.1 Structural Impulse Responses

Given $B_0$ and $u_t$, we immediately obtain $w_t = B_0 u_t$. Having identified the structural shocks $w_t$, our interest usually is not in the shocks themselves, however, but in the responses of each element of $y_t = (y_{1t}, \dots, y_{Kt})'$ to a one-time impulse in $w_t = (w_{1t}, \dots, w_{Kt})'$,

$$\frac{\partial y_{t+i}}{\partial w_t'} = \Theta_i, \quad i = 0, 1, 2, \dots, H,$$

where $\Theta_i$ is a $K \times K$ matrix. The elements of this matrix for given $i$ are denoted as

$$\theta_{jk,i} = \frac{\partial y_{j,t+i}}{\partial w_{kt}},$$

such that $\Theta_i = [\theta_{jk,i}]$. Usually the objective is to plot the responses of each variable to each structural shock over time. Since there are $K$ variables and $K$ structural shocks, there are $K^2$ impulse response functions, each of length $H + 1$, where $H$ is the maximum propagation horizon of the shocks.

A useful starting point for determining the structural impulse responses ($\theta_{jk,i}$) are the responses of $y_{t+i}$ to the reduced-form errors $u_t$. They can be obtained by considering the VAR(1) representation of the VAR($p$) process,

$$Y_t = \mathbf{A} Y_{t-1} + U_t, \tag{4.1.1}$$

where

$$Y_t \equiv \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, \quad \text{and } U_t \equiv \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

(see Chapter 2, Section 2.2.1). By successive substitution for $Y_{t-i}$, equation (4.1.1) can be written as

$$Y_{t+i} = \mathbf{A}^{i+1}Y_{t-1} + \sum_{j=0}^{i} \mathbf{A}^j U_{t+i-j}.$$

Left-multiplying this equation by $J \equiv [I_K, 0_{K\times K(p-1)}]$ yields

$$\begin{aligned} y_{t+i} &= J\mathbf{A}^{i+1}Y_{t-1} + \sum_{j=0}^{i} J\mathbf{A}^j U_{t+i-j} \\ &= J\mathbf{A}^{i+1}Y_{t-1} + \sum_{j=0}^{i} J\mathbf{A}^j J'JU_{t+i-j} \\ &= J\mathbf{A}^{i+1}Y_{t-1} + \sum_{j=0}^{i} J\mathbf{A}^j J'u_{t+i-j}. \end{aligned}$$

Thus, the response of the variable $j = 1, \ldots, K$ in the VAR($p$) system to a unit shock $u_{kt}$, $k = 1, \ldots, K$, $i$ periods ago, is given by

$$\underset{K\times K}{\Phi_i} = [\phi_{jk,i}] \equiv J\mathbf{A}^i J'.$$

The $\Phi_i$ are also sometimes referred to as responses to VAR forecast errors, as dynamic multipliers, or simply as reduced-form impulse responses.

If $y_t$ is covariance stationary, $y_t$ can be expressed as a weighted average of current and past shocks, with weights $\Phi_i$ that decline, the more distant a shock lies in the past. This multivariate MA representation is:

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \sum_{i=0}^{\infty} \Phi_i B_0^{-1} B_0 u_{t-i} = \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \tag{4.1.2}$$

where we made use of $w_{t-i} = B_0 u_{t-i}$ and defined $\Theta_i \equiv \Phi_i B_0^{-1}$. Under stationarity, it follows immediately that

$$\frac{\partial y_t}{\partial w'_{t-i}} = \frac{\partial y_{t+i}}{\partial w'_t} = \Theta_i.$$

The latter responses may be computed simply by post-multiplying $\Phi_i$, $i = 0, \ldots, H$, by $B_0^{-1}$:

$$\begin{aligned} \Theta_0 &= \Phi_0 B_0^{-1} = I_K B_0^{-1} = B_0^{-1} \\ \Theta_1 &= \Phi_1 B_0^{-1} \\ \Theta_2 &= \Phi_2 B_0^{-1} \\ &\vdots \end{aligned}$$

etc. If the VAR in question is not stable, the same approach to computing $\Phi_i$ and $\Theta_i$ will work, but the impulse responses may not approach zero for $i \to \infty$ and will no longer represent the coefficients of the structural MA representation. The $jk^{\text{th}}$ element of $\Theta_i$, denoted by $\theta_{jk,i}$, represents the response of variable $j$ to structural shock $k$ at horizon $i = 0, 1, \ldots, H$. There are $K^2$ impulse response functions

$$\theta_{jk,0}, \theta_{jk,1}, \ldots, \theta_{jk,H}$$

that trace the responses over time of each variable $j$ to each structural shock $k$.

It is also possible to construct linear combinations of structural impulse responses. For example, a VAR model may generate responses $\theta_{lk,i}$ for the nominal interest rate and $\theta_{mk,i}$ for the inflation rate, in which case the implied response of the real interest rate may be computed as $\theta_{lk,i} - \theta_{mk,i}$, $i = 0, \ldots, H$. Likewise, for example, we may infer from the response of the inflation rate, say $\Delta p_t$, the implied response of the log price level,

$$p_{t+h} = p_{t-1} + \Delta p_t + \Delta p_{t+1} + \cdots + \Delta p_{t+h},$$

by cumulating the responses of the inflation rate. In other words, the price level response to a shock in period $t$ at horizon $h$ is obtained as the sum

$$\sum_{i=0}^{h} \theta_{mk,i}$$

(see Lütkepohl 2005, section 2.3.2). Finally, it is also possible to compute nonlinear transformations of impulse response functions such as the half-life of responses, defined as the time that elapses until the response has fallen to half of the impact response in absolute value (see Kilian and Zha 2002).

A direct implication of the linearity of the VAR model is that responses to negative shocks are the mirror image of responses to positive shocks. Another implication of the linearity of the VAR model is that the magnitude of the structural shock does not matter in constructing impulse response functions, because rescaling the shock merely rescales the entire impulse response function.[1]

It is customary to choose $B_0^{-1}$ such that the structural shocks represent one standard deviation of the time series of structural shocks, because such a shock is viewed as a shock of typical magnitude, but nothing hinges on this convention. It is also important to keep in mind that structural shocks in general are unit free and cannot be expressed in terms of the units of measurement applied to the model variables. Only in special cases will structural shocks be associated with a particular model variable.

[1] If one does not like these implications, one has to turn to nonlinear VAR models instead (see Chapter 18).

To estimate the impulse responses in practice, all we need to do is replace the unknown parameters in the reduced-form VAR($p$) model by consistent estimates. Given estimates of the VAR parameters, $\widehat{A}_j$, $j = 1, \ldots, p$, and $\widehat{\Sigma}_u$ and the implied estimate of $B_0^{-1}$, we can construct $\widehat{\Phi}_i$ (and thus $\widehat{\Theta}_i$) recursively for $i = 0, \ldots, H$. For example, consider a VAR model proposed by Kilian and Park (2009) to shed light on the relationship between the global market for crude oil and the U.S. stock market. The model consists of the growth rate in global crude oil production, $\Delta prod_t$, a measure of the global business cycle in industrial commodity markets, $rea_t$, the real price of crude oil, $rpoil_t$, and U.S. real dividend growth, $\Delta rd_t$. Let $y_t = (\Delta prod_t, rea_t, rpoil_t, \Delta rd_t)'$. The objective is to decompose the reduced-form innovations into structural shocks representing oil supply shocks ($w_{1t}^{\text{oil supply}}$), shocks to the aggregate demand for all industrial commodities including crude oil ($w_{2t}^{\text{aggregate demand}}$), demand shocks that are specific to the oil market ($w_{3t}^{\text{oil-specific demand}}$), and a residual shock that captures all other determinants of U.S. real dividends ($w_{4t}^{\text{other}}$). By imposing a particular recursive ordering on $B_0^{-1}$ such that the elements above the diagonal are zero, the remaining elements of $B_0^{-1}$ can be recovered uniquely from $\Sigma_u$, as discussed in more detail in Chapter 8. Given an estimate of the $A_j$ and hence $\Phi_i$, $i = 0, \ldots, H$, knowledge of $B_0^{-1}$ allows us to estimate the corresponding structural impulse response matrices $\Theta_i$.

Now suppose that we are interested in quantifying the effects of oil demand and oil supply shocks on the level of U.S. real dividends (measured in percent deviations from the baseline). In other words, we are interested in the cumulative effects of oil supply and oil demand shocks on the fourth variable in the VAR model. These effects may be computed by cumulating the estimates of $\theta_{4,1,i}$, $\theta_{4,2,i}$, and $\theta_{4,3,i}$ for $i = 0, \ldots, H$. Figure 4.1 illustrates that a positive shock to global aggregate demand for all industrial commodities in the current month tend to raise U.S. real dividends by about 1% one year later, consistent with the view that a global demand boom has positive effects on the U.S. economy. In contrast, positive oil-specific demand shocks and negative oil supply shocks tend to lower real dividends by about 0.5% and 1%, respectively, consistent with the view that supply disruptions and other adverse events in the global oil market are detrimental to the U.S. economy. Figure 4.1 not only confirms prevailing views about the sign of these responses but allows us to quantify the effect of each shock.

## 4.2 Forecast Error Variance Decompositions

A second practically important question that a structural VAR model can answer is how much of the forecast error variance or prediction mean squared error (MSPE) of $y_{t+h}$ at horizon $h = 0, 1, \ldots, H$ is accounted for by each structural shock $w_{kt}$, $k = 1, \ldots, K$. In a stationary model, the limit of the

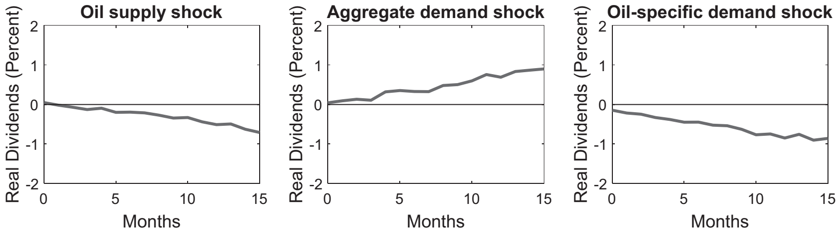


**Figure 4.1.** Point estimates of responses of U.S. real dividends to selected structural shocks
*Source:* Kilian and Park (2009).

forecast error variance decomposition, as $h \to \infty$, is the variance decomposition of $y_t$ because the forecast error covariance matrix or MSPE converges to the unconditional covariance matrix of $y_t$ (see Chapter 2). Thus, for stationary systems one may construct an MSPE decomposition for horizon infinity. In integrated systems the MSPE diverges when the forecast horizon goes to infinity, but the forecast error variance decomposition remains valid up to a finite maximum horizon of $H$. All we need to compute this decomposition is the $\Theta_i$ matrices, which we already computed for the structural impulse response analysis.

Recall from Chapter 2, Section 2.4.1, that for a VAR process the $h$-step ahead forecast error is

$$y_{t+h} - y_{t+h|t} = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i},$$

where $u_t = B_0^{-1} w_t$ allows us to replace $\Phi_i u_{t+h-i}$ by $\Theta_i w_{t+h-i}$. Hence, the MSPE at horizon $h$ is

$$\begin{aligned}
\mathrm{MSPE}(h) \equiv \mathbb{E}[(y_{t+h} - y_{t+h|t})(y_{t+h} - y_{t+h|t})'] &= \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' \\
&= \sum_{i=0}^{h-1} \Theta_i \underbrace{\Sigma_w}_{I_K} \Theta_i' \\
&= \sum_{i=0}^{h-1} \Theta_i \Theta_i'.
\end{aligned}$$

Let $\theta_{kj,h}$ be the $kj^{\text{th}}$ element of $\Theta_h$. Then the contribution of shock $j$ to the MSPE of $y_{kt}$, $k = 1, \dots, K$, at horizon $h$ is

$$\mathrm{MSPE}_j^k(h) = \theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2$$

Table 4.1. *Forecast Error Variance Decomposition for U.S. Real Dividend Growth*

| | Percent of $h$-Step Ahead Forecast Error Variance Explained by: | | | |
|---|---|---|---|---|
| Horizon | Oil supply shock | Aggregate demand shock | Oil-specific demand shock | Residual shock |
| 1 | 0.2 | 0.2 | 1.7 | 98.0 |
| 2 | 0.6 | 0.4 | 2.1 | 97.0 |
| 3 | 0.8 | 0.5 | 2.1 | 96.6 |
| 12 | 2.8 | 6.8 | 4.5 | 85.8 |
| $\infty$ | 6.6 | 8.4 | 7.9 | 77.1 |

*Source:* Kilian and Park (2009).

and the total MSPE of $y_{kt}$, $k = 1, \ldots, K$, at horizon $h$ is:

$$\text{MSPE}^k(h) = \sum_{j=1}^{K} \text{MSPE}_j^k(h) = \sum_{j=1}^{K} \left(\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2\right).$$

Dividing

$$\text{MSPE}^k(h) = \sum_{j=1}^{K} \text{MSPE}_j^k(h)$$

by $\text{MSPE}^k(h)$ yields the following decomposition for given $h$ and $k$:

$$1 = \frac{\text{MSPE}_1^k(h)}{\text{MSPE}^k(h)} + \frac{\text{MSPE}_2^k(h)}{\text{MSPE}^k(h)} + \cdots + \frac{\text{MSPE}_K^k(h)}{\text{MSPE}^k(h)},$$

where each ratio gives the fraction of the contribution of the $j^{\text{th}}$ shock to the MSPE($h$) of variable $k$ for $j = 1, \ldots, K$. In other words, $\text{MSPE}_j^k(h)/\text{MSPE}^k(h)$ is the fraction of the contribution of shock $j$ to the forecast error variance of variable $k$. By multiplying the fractions by 100 we obtain percentages.

The use of forecast error variance decompositions again is best illustrated by example. Returning to the four-variable VAR model considered by Kilian and Park (2009), an obvious question raised by their analysis is to what extent variability in U.S. real dividend growth can be explained by oil demand and oil supply shocks. This question may be answered based on a forecast error variance decomposition for U.S. real dividend growth. Kilian and Park focus on horizons of 1, 2, 3, and 12 months. Since real dividend growth ($\Delta rd_t$) is $I(0)$, a forecast horizon of $h = \infty$ is also considered. The results are presented in Table 4.1.

Ignoring rounding error, the entries in each row of the table sum to 100% by construction. The entries for horizon $\infty$ represent the variance decomposition of U.S. real dividend growth. In practice, we can approximate $\infty$ by a very large number. This number is determined by showing that further increments to the horizon do not change the results up to the desired degree of accuracy.

In studying forecast error variance decompositions, one often is interested in the patterns across horizons. In this example we learn that the oil supply shock and the two oil demand shocks combined account for only 2% of the MSPE of U.S. real dividend growth at the one-month horizon, but that their explanatory power increases to 23% in the long run. This is evidence that the relationship between the global oil market and the U.S. stock market is weak at best. One may also be interested in the relative contribution of different shocks at a given horizon. For example, whereas at the one-month horizon oil specific demand shocks are much more important than oil supply shocks or aggregate demand shocks in explaining the forecast error variance of real dividend growth, each oil demand and oil supply shock accounts for about the same share of the unconditional variance.

## 4.3 Historical Decompositions

Structural forecast error variance decompositions and structural impulse response functions describe the average movements in the data. They represent unconditional expectations. Sometimes we are interested instead in quantifying how much a given structural shock explains of the historically observed fluctuations in the VAR variables. In other words, we would like to know the cumulative effect of a given structural shock on each variable at every given point in time. For example, we may not be interested in the average contribution of monetary policy shocks to the variability of real GDP growth over the last decades, but in the question of whether monetary policy shocks caused the 1982 recession.

Such historical decompositions may be computed from covariance stationary VAR models as follows. Suppose that we have data from 1 to $t$. Then, for any $t$,

$$y_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s} + \sum_{s=t}^{\infty} \Theta_s w_{t-s}.$$

In other words, the value of $y_t$ depends on shocks $w_1, \ldots, w_t$ that can be estimated and shocks that predate the start of the sample at $t = 1$ and hence cannot be estimated. Given the fact that the MA coefficients die out, as we move further into the past, the second term (corresponding to the presample period) will have a steadily diminishing effect on $y_t$ as $t$ increases. Even after dropping the second term, it will be the case that

$$y_t \approx \sum_{s=0}^{t-1} \Theta_s w_{t-s}, \tag{4.3.1}$$

except for a number of periods early in the sample. We denote this approximation by

$$\widehat{y}_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s}. \tag{4.3.2}$$

We start by plotting $\widehat{y}_t$ and the (suitably demeaned and detrended) actual data $y_t$ in the same plot. We discard the initial observations (also known as transients), for which the two series do not effectively coincide. How many periods it takes for the approximation to work well depends on the dominant root of the VAR process (defined as the root closest to the unit circle). The closer the dominant root is to unity, the more persistent is the effect of the shocks on $y_t$ and the longer the period of transition. For the remaining sample period, we can decompose the sum in (4.3.2) to isolate the cumulative contribution of each shock to each element of $\widehat{y}_t$, as discussed next.

***Historical Decompositions as Time Series Plots.*** In practice, the construction of historical decompositions involves three simple steps. Step 1 is to compute the structural MA coefficient matrices $\Theta_0, \ldots, \Theta_{T-1}$. Step 2 is to compute the structural shocks $w_t = B_0 u_t, t = 1, \ldots, T$. Step 3 is to match up each structural shock, say shock $j$, with the appropriate impulse response weight, as required by the structural moving average representation, to form $T \times 1$ vectors of fitted values for variable $k$, denoted $\widehat{y}_k^{(j)}$ for $j = 1, \ldots, K$.

For example, let $K = 5$ and suppose that we are interested in the cumulative effect of each of the five structural shocks on the fourth variable of the VAR system. In that case, we compute the following weighted sums for $t = 1, \ldots, T$:

$$\widehat{y}_{4t}^{(1)} = \sum_{i=0}^{t-1} \theta_{41,i} w_{1,t-i},$$

$$\widehat{y}_{4t}^{(2)} = \sum_{i=0}^{t-1} \theta_{42,i} w_{2,t-i},$$

$$\widehat{y}_{4t}^{(3)} = \sum_{i=0}^{t-1} \theta_{43,i} w_{3,t-i},$$

$$\widehat{y}_{4t}^{(4)} = \sum_{i=0}^{t-1} \theta_{44,i} w_{4,t-i},$$

$$\widehat{y}_{4t}^{(5)} = \sum_{i=0}^{t-1} \theta_{45,i} w_{5,t-i},$$

where $\theta_{jk,i}$ denotes the response of variable $j$ to shock $k$ at horizon $i$ and $w_{k,t}$ is the $k^{\text{th}}$ structural shock at time $t$. Each vector $\widehat{y}_4^{(j)} = (\widehat{y}_{41}^{(j)}, \ldots, \widehat{y}_{4T}^{(j)})'$ shows the cumulative contribution of shock $j$ on the fourth variable in the VAR model over time. By construction, the value for $\widehat{y}_{4t}$ is obtained as the sum

$$\widehat{y}_{4t} = \sum_{j=1}^{K} \widehat{y}_{4t}^{(j)}.$$

In practice, historical decompositions are computed by replacing the unknown quantities $\theta_{jk,i}$ and $w_t$ by the usual estimates.

It is important to remember that historical decompositions involve an approximation error. This approximation error arises because we truncate the moving average representation. For example, $y_{k1}$ depends on the structural shocks at date 1 as well as the infinite history of structural shocks. With much of the history of shocks unobserved, the approximation is bound to be poor initially. As we recursively update $\widehat{y}_{kt}$, however, more and more of the recent structural shocks that receive high impulse response weights are captured, and the weights of earlier unobserved shocks decline. Thus, $\widehat{y}_{kt}$ approaches $y_{kt}$. How fast this convergence takes place depends on the persistence of $y_{kt}$.

The best way to determine the point $T^*$ beyond which the historical decomposition is reasonably accurate is to plot both $\widehat{y}_{kt}$ and $y_{kt}$ against time. If we plot the historical decomposition along with actual data, we must first remove all deterministic components in the actual data, because the data generated from the structural MA representation (4.1.2) are zero mean by construction. All data predating the point of convergence must be discarded. Only $\widehat{y}_{kt}^{(j)}$, $j = 1, \ldots, K$, for $t = T^*, T^* + 1, \ldots, T$ can be used for the analysis of the historical decomposition. When the sample is short and the data are persistent, this fact may preclude the use of this tool.

The way to interpret historical decompositions is that each $\widehat{y}_{kt}^{(j)}$, $j = 1, \ldots, K$, shows the cumulative effect of shock $j$ on $y_{kt}$ up to this point in time. This cumulative effect may be positive or negative relative to the benchmark of zero, which represents the mean value of the data. Each time series $\widehat{y}_{kT^*}^{(j)}, \ldots, \widehat{y}_{kT}^{(j)}$, $j = 1, \ldots, K$, shows how our approximate variable $\widehat{y}_{kt}$ would have evolved if all other structural shocks had been turned off. Plotting each time series $\widehat{y}_{kT^*}^{(j)}, \ldots, \widehat{y}_{kT}^{(j)}$, $j = 1, \ldots, K$, against the actual data helps assess which structural shock(s) alone or in combination account for the fluctuations in $\widehat{y}_{kT^*}, \ldots, \widehat{y}_{kT}$ during specific historical episodes of interest. By construction, the sum

$$\sum_{j=1}^{K} \widehat{y}_{kt}^{(j)}$$

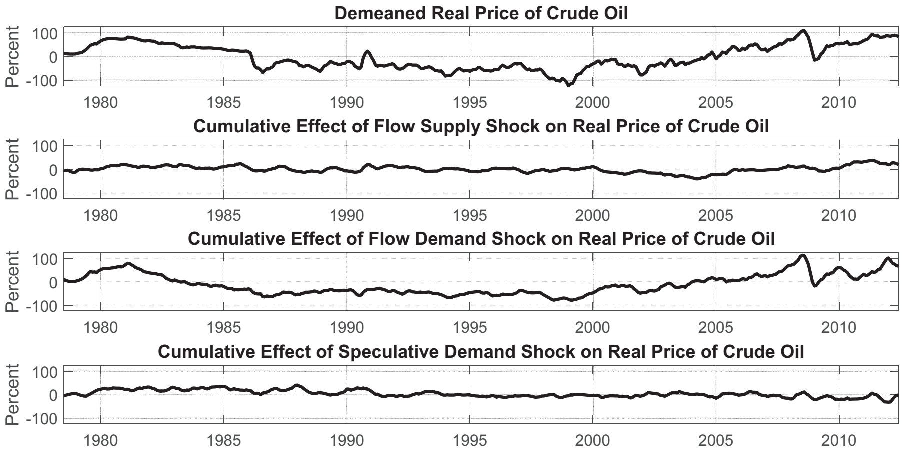


**Figure 4.2.** Historical decomposition of the real price of crude oil in percent deviations from the mean.
*Source:* Kilian and Lee (2014).

must equal the demeaned actual data except for model estimation error and approximation error arising from the truncation of the infinite sum. Because the structural MA representation is based on a parametric estimate of the mean of the data, defined as the $k^{\text{th}}$ element of $(I_K - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1}\widehat{\nu}$, and that estimate in finite samples may differ from the sample average of the data, defined as the $k^{\text{th}}$ element of $\bar{y}$, $\widehat{y}_{kt}$ may in practice differ slightly from the demeaned $y_{kt}$ by a constant, even after disregarding the transients. When plotting both $\widehat{y}_{kt}$ and the demeaned $y_{kt}$ against time, this discrepancy may be removed by demeaning $\widehat{y}_{kt}$.

Figure 4.2 illustrates the use of historical decompositions in understanding the evolution of the real price of oil from the late 1970s to early 2012. The example is based on a global oil market model studied in Kilian and Lee (2014) which is discussed in more detail in Chapter 13. This structural model attributes variation in the real price of oil to shocks to the flow supply of oil, shocks to the flow demand for oil, a speculative oil demand shock, and a residual shock designed to capture various idiosyncratic shocks.

Suppose that we are interested in understanding what happened in the global oil market during a particular historical episode. For example, we may be interested in explaining the surge in the real price of oil between 2003 and mid-2008, keeping in mind that different historical episodes may be explained by different combinations of structural shocks. Figure 4.2 focuses on the role of the structural shocks that have an explicit economic interpretation under the maintained assumption that the real price of oil is an $I(0)$ time series during the estimation period. It shows that much of this surge (as well as the collapse of the real price of oil in late 2008 and its recovery since then) must be attributed to the effects of flow demand shocks. Neither flow supply shocks nor speculative demand shocks are able to explain the surge in the real price of oil during this period. This result could not have been inferred from the structural impulse responses that trace out the average effect of a hypothetical one-time structural shock, or from forecast error variance decompositions that measure the extent to which a structural shock explains the variability of a variable on average.

The importance of conducting historical decompositions is not always appreciated in empirical macroeconomics. An occasional mistake in the business cycle literature is to view evidence of large and persistent impulse responses of real output to a structural shock as evidence of this shock's ability to explain the business cycle. This conclusion is unwarranted because impulse responses trace out the response to a one-time positive shock only, whereas business cycle variation in real output is driven by a sequence of shocks of different magnitude and signs. It is not uncommon for the effects of a positive shock in one period to be eclipsed by negative shocks in subsequent periods. Only the historical decomposition allows us to assess the cumulative effect of these shocks on the business cycle and the relative importance of different shocks in explaining particular recessions or expansions.

***Extensions to Integrated Time Series.*** Unlike structural impulse responses, historical decompositions are designed for stationary VAR variables. They cannot be applied to integrated or cointegrated variables in levels without modifications because they rely on the existence of a stationary MA representation of the DGP. A case in point is a VAR model containing real GDP. Suppose we think of this time series as containing a unit root. One way of proceeding would be to express this time series in growth rates and to apply the historical decomposition to the growth rate. The disadvantage of that approach is that it involves a loss of information. We cannot learn anything about the level of real GDP predicted by the model because that information is lost when converting the data to growth rates. We can, however, quantify the ability of a given shock to explain the cumulative change in real GDP since a given point in time.

If the possibly integrated real GDP variable is included in levels in the VAR model, as is common in applied work, it is not possible to construct a historical decomposition for real GDP at all. Sometimes VAR users report the results of an exercise in which, starting at some date $T^*$, all subsequent realizations of shock $j$ are replaced by zero. The difference between the observed values of real GDP after $T^*$ and those obtained under this conditional counterfactual are then taken as the cumulative effect of this structural shock since $T^*$. While this approach has merit, if we are interested in the incremental role played by shock $j$ during a particular episode in the data, it differs from a historical

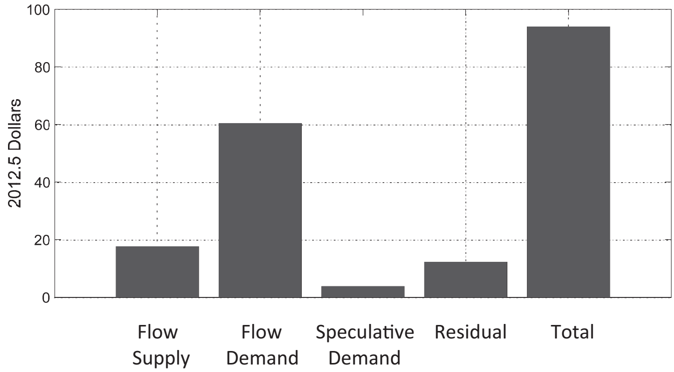


**Figure 4.3.** Contribution of each structural shock to the cumulative change in the real price of oil from January 2003 to June 2008.
*Source:* Adapted from Kilian and Lee (2014).

decomposition in that it ignores the cumulative effects of shock $j$ prior to $T^*$ on the subsequent real GDP realizations.

***Historical Decompositions as Bar Charts.*** Kilian and Lee (2014) proposed two alternative simpler formats for presenting the information conveyed by historical decompositions. The first format allows one to express the cumulative change in a stationary variable explained by each structural shock in the form of a bar chart. To continue with the earlier example, let $\widehat{y}_{4T_1}^{(j)}$ denote the cumulative contribution of shock $j$ to variable $y_{4t}$ at date $T_1$ and $\widehat{y}_{4T_2}^{(j)}$ the corresponding result for date $T_2 > T_1$. Then the cumulative change in $y_{4t}$ between dates $T_1$ and $T_2$ contributed by shock $j$ is approximated by:

$$\widehat{y}_{4T_2}^{(j)} - \widehat{y}_{4T_1}^{(j)},$$

with the total change in $y_{4t}$ measured as:

$$\sum_{j=1}^{5} \left( \widehat{y}_{4T_2}^{(j)} - \widehat{y}_{4T_1}^{(j)} \right) = \widehat{y}_{4T_2} - \widehat{y}_{4T_1} \approx y_{4T_2} - y_{4T_1}.$$

This allows us to provide a quick summary of the evidence for any subperiod of interest. Kilian and Lee (2014) use this tool to summarize the determinants of the surge in the real price of oil between January 2003 and June 2008 on the one hand and since the peak of the real price of oil on the other. Figure 4.3 shows an example of such a bar chart. The bar on the right shows that between 2003 and mid-2008 the price of oil in total increased by 95 dollars in real terms. The four bars on the left indicate how much of this 95-dollar increase must be attributed to each of the four structural shocks. They show that 18 dollars of this increase are explained by the flow supply shock, 61 dollars by the flow demand shock, 4 dollars by the speculative demand shock, and 12 dollars by the residual shock, providing evidence against the hypothesis that speculative demand was a quantitatively important determinant of the surge.

***Historical Decompositions as Counterfactuals.*** An alternative way of assessing the cumulative contribution of each shock to $y_{kt}$, also proposed in Kilian and Lee (2014), is to construct the counterfactual

$$y_{kt} - \widehat{y}_{kt}^{(j)},$$

where $y_{kt}$ denotes the $k^{\text{th}}$ actual time series variable (including any deterministic means or trends) and $\widehat{y}_{kt}^{(j)}$ is the cumulative contribution of shock $j$ to the evolution of variable $k$ up to date $t$, as defined earlier. This representation avoids the impossible task of having to attribute the deterministic component of $y_{kt}$ to individual shocks $j$. The counterfactual series indicates how the variable of interest would have evolved had one been able to replace all realizations of shock $j$ by zero while preserving the remaining structural shocks in the model. If the counterfactual exceeds $y_{kt}$, this means that structural shock $j$ lowered $y_{kt}$. A counterfactual below the observed $y_{kt}$ means that the shock in question raised $y_{kt}$. The vertical distance between the actual $y_{kt}$ and the counterfactual tells us how much shock $j$ affected $y_{kt}$ at this point in time.

It is useful to contrast this approach with the earlier approach of constructing cumulative increases in $\widehat{y}_{kt}^{(j)}$ and $y_{kt}$. That approach focused on changes in the variables over time explained by a given structural shock rather than the component of $y_{kt}$ at a given point in time driven by that structural shock. To move from a plot of the counterfactual to the cumulative increase measure, one would have to compare the difference between the counterfactual and $y_{kt}$ on the first and on the last date of the counterfactual and construct the rate of change over time in this difference. Thus, these representations are mutually consistent but focus on different aspects of the same data.

Figure 4.4 provides an example of some historical counterfactuals based on structural VAR models. It characterizes the evolution of the real price of oil from January 2003 to June 2008 based on the global oil market model of Kilian and Lee (2014), again maintaining the assumption that the real price of oil is $I(0)$. The counterfactuals show how the real price of oil would have evolved, had one been able to replace all realizations of a given structural shock by zero while preserving the remaining structural shocks in the model. The first panel shows, for example, that in the absence of flow supply shocks, the level of the real price of oil would have been higher between 2003 and mid-2005. In other words, positive flow supply shocks helped keep the real price of oil down. For the remainder of the sample, flow supply shocks played only a minor role. The second panel shows that, in the absence of flow demand shocks, the real price of oil would have been somewhat higher in 2003, but much lower starting

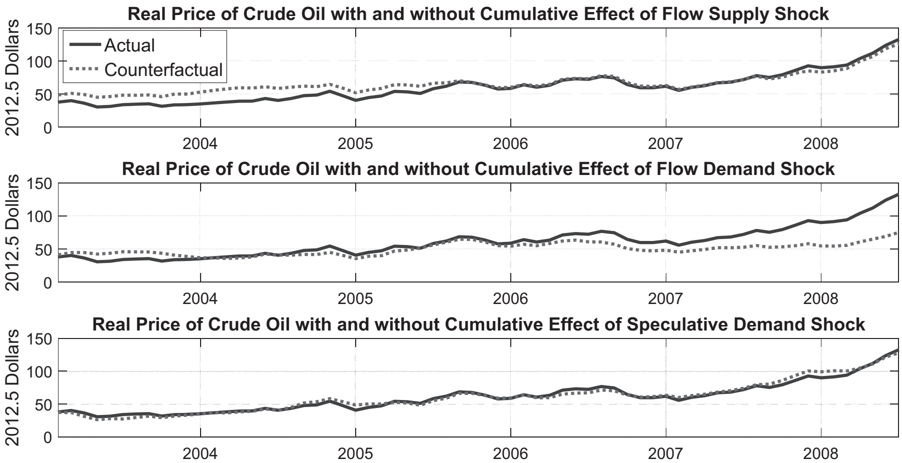


**Figure 4.4.** Historical counterfactuals for the real price of crude oil from January 2003 to June 2008
*Source:* Kilian and Lee (2014).

in 2006. Finally, the third panel demonstrates that speculative demand shocks made little difference overall. In fact, they slightly lowered the real price of oil in late 2007 and early 2008, according to this model.

## 4.4 Forecast Scenarios

The objective of forecast scenarios is to assess the sensitivity of reduced-form VAR forecasts to hypothetical future events. Constructing such forecast scenarios requires a structural VAR model, the reduced-form representation of which generates accurate out-of-sample forecasts. It is important to keep in mind that the objective of constructing forecast scenarios is not to improve the accuracy of the baseline reduced-form VAR forecast. Indeed, that forecast by construction already provides the best possible out-of-sample prediction from a given forecasting model.

Unlike a conventional forecast, forecast scenarios are not designed to characterize the most likely outcomes given past observations, but to characterize the risks associated with unlikely possible outcomes in the future in the form of a hypothetical “what-if” question. For example, we may ask how the occurrence of a certain future event, no matter how unlikely, would affect the future values of the variable that is being forecast. This is a question often raised by policymakers.

The simplest form of a forecast scenario is a conditional point forecast. More generally, the sensitivity of a forecast to alternative assumptions about future structural shocks or observables may be captured by probability-weighted conditional forecast densities and summarized by formal risk measures (see Baumeister and Kilian 2014). Risk analysis along these lines allows users of VAR forecasts to explore by how much and at what forecast horizon downside and upside risks change as a function of the probability weights attached to different scenarios. In conjunction with historical decompositions, this approach helps satisfy the needs of policymakers not only for accurate out-of-sample forecasts but for a coherent economic interpretation of both historical data and forecasts.

There are two closely related but distinct approaches to the construction of conditional forecasts from structural VAR models. One conditions on sequences of future structural shocks. This type of conditional forecast was developed by Baumeister and Kilian (2014). The other approach relies on forecasts conditional on the future path of observables. The latter approach was proposed by Waggoner and Zha (1999).

### *4.4.1 Conditional Forecasts Expressed in Terms of Sequences of Structural Shocks*

Following Baumeister and Kilian (2014), forecast scenarios are constructed from the structural moving average representation of the VAR model. We first iterate that representation forward to obtain:

$$y_{t+h} = \sum_{i=0}^{\infty} \Theta_i w_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i},$$

where $y_{t+h}$ denotes the dependent variable $h$ periods in the future, which can be written as the sum of two terms. The second term captures the cumulative effects on $y_{t+h}$ of all structural shocks that already occurred between $-\infty$ and $t$. This term is predetermined at $t$. The first term captures the cumulative effect on $y_{t+h}$ of all structural shocks that have yet to occur between $t+1$ and $t+h$. When computing forecasts, the conventional approach is to set this first term to zero in the forecast equation because each structural shock in the model is unconditionally zero in expectation. It follows that the second term represents the unconditional model-based expectation (or forecast of $y_{t+h}$ at date $t$),

$$y_{t+h|t} = \sum_{i=h}^{\infty} \Theta_i w_{t+h-i},$$

which is obtained by setting the future shocks, $w_{t+1}, \ldots, w_{t+h}$, to zero. This forecast is identical by construction to the date $t$ forecast that would be obtained from iterating forward the autoregressive reduced-form representation of the VAR model.

Feeding in a sequence of nonzero future structural shocks and conditioning on them in computing the forecast, in contrast, provides a conditional point forecast. Such forecast scenarios may be based on purely hypothetical shock sequences, as long as the shock sequence considered is within the range of historical experience. This caveat is reminiscent of the discussion of modest policy interventions in econometric models in Leeper and Zha (2003). Shocks that are unusually large by historical standards not only may induce changes in behavior of the type emphasized by Lucas (1976), but there also would be reason to doubt the adequacy of standard linear model approximations when considering shock sequences involving extremely large shocks.

Forecast scenarios may also be based on sequences of structural shocks that occurred during selected episodes in the past. A historical decomposition can help identify such shock sequences. For example, Baumeister and Kilian (2014) identify the sequence of structural shocks in the oil market associated with events such as the Asian crisis of 1997 or the collapse of oil prices in 2008, and explore the consequence of a recurrence of these events for forecasts of the real price of oil. When postulating shock sequences directly based on historical precedent, the Lucas (1976) critique does not apply by construction, given the maintained assumption of a well-specified structural VAR model.

A final caveat is that forecast scenarios by construction involve sequences of unanticipated shocks. This does not mean that one cannot construct forecast scenarios involving anticipation, but that doing so requires a structural VAR model that allows for anticipation. An example is provided in Baumeister and Kilian (2014). For further discussion, see Chapter 17.

***Conditional Point Forecasts.*** It is convenient to normalize all conditional forecasts relative to the baseline forecast from the structural model obtained by setting all future structural shocks to zero, which eliminates the dependence of the forecast scenario on $t$. The plot of this normalized conditional forecast represents the upward or downward adjustments of the baseline forecast that would be required if a given hypothetical scenario were to occur. More formally, for a vector of future structural shocks $\left\{w_{t+1}^{\text{scenario}}, w_{t+2}^{\text{scenario}}, \ldots, w_{t+h}^{\text{scenario}}\right\}$ with nonzero mean, the revision required in the baseline forecast of $y_{t+h}$, $h = 1, 2, \ldots,$ in expectation would be:

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i} \,\middle|\, \{w_{t+h-i} = w_{t+h-i}^{\text{scenario}}\}_{i=0}^{h-1}, \Omega_t\right) \\
&- \mathbb{E}\left(\sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i} \,\middle|\, \{w_{t+h-i} = w_{t+h-i}^{\text{baseline}}\}_{i=0}^{h-1}, \Omega_t\right) \\
&= \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}^{\text{scenario}},
\end{aligned}
$$

where $\Omega_t = \{w_{t+h-i}, i = h, h+1, \ldots\}$ denotes the information set of the forecaster at date $t$ and $\sum_{i=0}^{h-1} \Theta_i w_{t+h-1}^{\text{baseline}} = 0$ because $w_{t+h-i}^{\text{baseline}} = 0$ for $i = 0, \ldots, h-1$. This approach is similar to the construction of impulse response functions. The key difference is that an impulse involves a one-time structural shock $w_{t+1}^{\text{scenario}} \neq 0$ followed by $w_{t+i}^{\text{scenario}} = 0$ for $i = 2, 3, \ldots, h$, whereas forecast scenarios tend to involve sequences of nonzero structural shocks extending over several periods. Once the forecast scenarios have been expressed in percent deviations from the baseline forecast, the scenario forecasts are constructed by scaling the baseline reduced-form VAR forecast. This simply involves adding the deviation from the baseline forecast to the baseline forecast such that the scenario forecast reduces to

$$y_{t+h|t} + \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}^{\text{scenario}}.$$

In practice, forecast scenarios are constructed by replacing $\Theta_i$ by a consistent estimate. An important difference between conditional and unconditional forecasting methods is that, as with impulse response analysis, in generating conditional forecasts, the objective is to identify to the best of our ability the true dynamic effects of a sequence of structural shocks, which necessitates the use of fully revised data. Accordingly, the estimates of $\Theta_i$ should be based on fully revised data, even if the unconditional forecasts are based on real-time data. Moreover, given the linearity of the VAR model, forecast scenarios (like impulse response functions) are time invariant. In other words, a given hypothetical sequence of shocks will cause the same revisions of the baseline forecast at each point in time. This also means that deviations from the baseline in practice do not necessarily have to be recomputed every month, except to the extent that a longer sample offers efficiency gains in estimating the structural model.

An example of the use of forecast scenarios is shown in Figure 4.5. Based on a model of the global oil market similar to that used in Kilian and Lee (2014), Baumeister and Kilian (2014) investigate a wide range of real-time forecast scenarios for the real price of crude oil, including a return of Iraqi oil production to full capacity, a supply disruption in Libya, a strong recovery of the global economy, a financial meltdown similar to the collapse of Lehman Brothers, and two contagion scenarios in which expectations of rising oil prices are triggered by political events in the Middle East. Some of these scenarios are based on historical precedent, while others are purely hypothetical. All scenarios involve sequences of structural shocks within the range of historical experience. Figure 4.5 shows how the forecast of the real price of oil would deviate from the baseline real-time VAR forecast as of December 2010 if one were willing to condition on each one of these events occurring in isolation. Such evidence allows policymakers to gauge the potential effects of unlikely but high-impact events on the real price of crude oil.

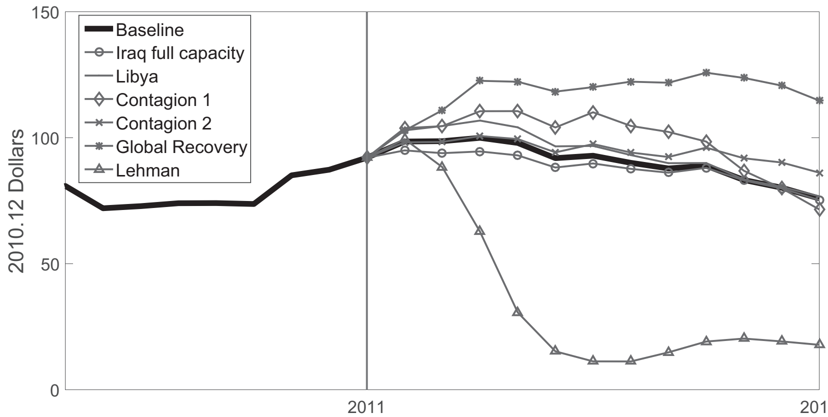


**Figure 4.5.** Selected real-time forecast scenarios for the real price of crude oil as of December 2010.
*Source:* Adapted from Baumeister and Kilian (2014).

***Conditional Forecast Densities.*** So far our analysis has focused on conditional point forecasts and has abstracted from the uncertainty surrounding these forecasts. A more formal evaluation of the risks inherent in forecasts requires the construction of conditional density forecasts. We postulate that the scenario errors, $w_{t+1}^{\text{scenario}}, w_{t+2}^{\text{scenario}}, \ldots, w_{t+h}^{\text{scenario}}$, have the same distribution as the baseline errors, $w_{t+1}^{\text{baseline}}, w_{t+2}^{\text{baseline}}, \ldots, w_{t+h}^{\text{baseline}}$, during the forecast period except that the mean of this distribution has changed to the alternative shock sequence associated with the scenario. This allows one to construct the conditional forecast density for a given forecast scenario by shifting the unconditional forecast density and centering it on the conditional point forecast implied by the forecast scenario.

The unconditional forecast density may be obtained by bootstrap methods. For example, if the VAR errors are iid, we can generate sequences of random shocks, $w_{t+1}, w_{t+2}, \ldots, w_{t+h}$, by drawing with replacement from the set of model residuals, allowing us to simulate the future path of the data conditional on the information set at date $t$ and the estimated VAR model parameters. Having simulated a large number of realizations of the sequence $y_{k,t+1}, y_{k,t+2}, \ldots, y_{k,t+h}$, an estimate of the forecast density for $y_{k,t+h}$ at a given forecast horizon $h$ may be constructed using a Gaussian kernel density smoother. An example is the baseline 1-year ahead forecast density shown in Figure 4.6. For a discussion of more sophisticated boostrap methods for predictive inference that allow for estimation uncertainty in the VAR model parameters the reader is referred to Chapter 12.

The conditional forecast density is all we need when considering one forecast scenario at a time. Alternatively, several forecast scenarios may be combined into one conditional point or density forecast if one is willing to

 https://doi.org/10.1017/9781108164818.005

attach probability weights to each individual scenario. These probabilities reflect how much weight the end user wishes to attach to each one of a set of hypothetical events. They allow one to explore the consequences of conditioning on combinations of events occurring in expectation with given probability weights. One example would be an analysis of two scenarios in conjunction. In this case, it can be useful to consider the implications of the respective probabilities of only one of these two scenarios occurring, or of both occurring in conjunction, instead of postulating that only one of the three possible scenarios occurs. Alternatively, one may wish to explore how the increasing probability of an extreme event affects the forecast compared with the baseline forecast.

The construction of probability-weighted scenarios is a natural generalization of the idea of conditioning on one scenario at a time, which amounts to assigning probability one to this scenario and probability zero to all other scenarios. Conditioning on hypothetical events, with the mean of the distribution expressed as a probability-weighted average, serves to explore the sensitivity of the forecast to these events. A useful tool in constructing the probability weights is a Venn diagram that illustrates the relationship between alternative scenarios. The ultimate objective of constructing probability-weighted conditional density forecasts is to summarize the results in the form of formal risk measures. This allows the policymaker to explore how changes in the probability weights attached to different scenarios affect the upside and downside risks of the forecast, providing insights into the determinants of the baseline forecast.

Figure 4.6 illustrates how alternative scenarios may affect the predictive density generated from the baseline VAR forecasting model for the real price of oil. All results take the information set available to forecasters in December 2010 as given. The baseline density implies substantial uncertainty about the real price of oil one year into the future. Figure 4.6 also examines how the density changes under three alternative scenarios that differ primarily in how optimistic the user is about the future state of the global economy. The more optimistic the user is about the future state of the global economy, the higher the probability weight on the global recovery scenario compared with the other scenarios discussed earlier. More optimism translates to higher demand for crude oil in expectation and hence shifts the oil price density to the right without affecting its shape. Increased pessimism shifts the density to the left. Figure 4.6 shows that the baseline predictive density almost coincides with the density under the moderately pessimistic scenario, but expectations of a global recovery may shift the predictive density substantially to the right.

The extent to which predictive densities evolve under alternative scenarios can be difficult to convey to policymakers. It is therefore useful to summarize these results based on formal measures of oil price risks, building on the work

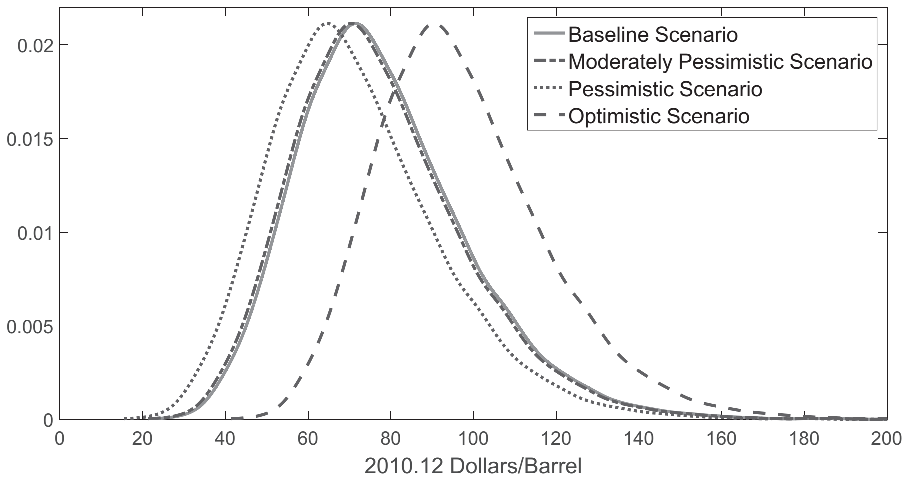


**Figure 4.6.** Real-time probability weighted 1-year ahead density forecasts as of December 2010 under different scenarios about the future state of the global economy. *Source:* Adapted from Baumeister and Kilian (2014).

of Kilian and Manganelli (2007). Let $R_{t+h}$ denote the real price of oil $h$ months from now measured in December 2010 dollars and $F$ the predictive distribution of $R_{t+h}$. For expository purposes, consider the event of $R_{t+h}$ exceeding an upper threshold of \$100. Then, under conventional assumptions about the degree of risk aversion, the upside risk in the oil price forecast, defined as

$$\int_{100}^{\infty} (R_{t+h} - 100)\, dF,$$

can be expressed as

$$\mathbb{E}(R_{t+h} - 100 | R_{t+h} > 100) \times \mathbb{P}(R_{t+h} > 100),$$

where the first term of the expression is referred to as the tail-conditional expectation and the second term as the tail probability. The first column of Table 4.2 reports the tail probability of the real price of oil exceeding \$100; the second column shows the tail conditional expectation (or expected excess) which tells us by how much one would expect to exceed \$100 conditional on exceeding that threshold, and the last column reports the weighted expected excess defined as the product of the first two terms. Similar risk measures could also be computed for the risk of the real price of oil falling below \$80, for example, depending on the event the user wishes to guard against. It may also be of interest to study the balance of the upside and downside risks.

Table 4.2. *Real-Time Risk Measures as of December 2010*

| Scenario | $h$ | $\mathbb{P}(R_{t+h} > 100)$ | $\mathbb{E}(R_{t+h} - 100 \mid R_{t+h} > 100)$ | $\mathbb{E}(R_{t+h} - 100 \mid R_{t+h} > 100) \times \mathbb{P}(R_{t+h} > 100)$ |
|---|---|---|---|---|
| Baseline | 3 | 0.50 | 11.03 | 5.56 |
| | 6 | 0.36 | 15.25 | 5.42 |
| | 12 | 0.14 | 15.40 | 2.22 |
| Moderately | 3 | 0.56 | 11.51 | 6.42 |
| Pessimistic | 6 | 0.31 | 14.81 | 4.60 |
| | 12 | 0.14 | 15.29 | 2.08 |
| Pessimistic | 3 | 0.44 | 10.49 | 4.57 |
| | 6 | 0.19 | 13.69 | 2.59 |
| | 12 | 0.10 | 14.76 | 1.41 |
| Optimistic | 3 | 0.86 | 16.78 | 14.48 |
| | 6 | 0.63 | 18.41 | 11.60 |
| | 12 | 0.40 | 17.82 | 7.10 |

*Source:* Adapted from Baumeister and Kilian (2014). The more optimistic we are about the future state of the global economy, the larger the upside risk for the real price of oil.

Table 4.2 shows that under the baseline scenario the probability of the real price of oil exceeding \$100 is 50% at the 3-month horizon, but declines to 36% at the 6-month horizon and to 14% at the 12-month horizon. The expected excess is about \$11, but rises to \$15 at longer horizons. Overall, the upside risk as measured by the weighted expected excess declines with the horizon. Changes to the baseline scenario are reflected in somewhat different risk estimates. Conditioning on the optimistic scenario, for example, increases the probability of the real price of oil surpassing \$100 to 86% at the 3-month horizon, coupled with an expected excess of \$17.

### *4.4.2 Conditional Forecasts Expressed in Terms of Sequences of Observables*

In closely related work, Waggoner and Zha (1999) proposed the construction of forecasts conditional on the path of the observables. This approach makes sense when modeling the path of a policy instrument such as the short-term interest rate under the premise that the central bank can control the policy instrument by executing a sequence of interest rate shocks. In that case, forecast scenarios are based on the sequence of monetary policy shocks implied by the hypothesized path of the interest rate. Waggoner and Zha (1999) show how to construct conditional point forecasts and conditional forecast densities in the context of a monetary policy VAR model using a Bayesian approach to estimation and inference (see Chapter 5).

The key difference between Waggoner and Zha's forecasts conditional on the interest rate and Baumeister and Kilian's forecast scenarios for the real price of oil is that in oil markets none of the many structural shocks affecting the variable of interest can be controlled by the policymaker. Moreover, there would be no point in constructing scenarios in terms of a prespecified path of the real price of oil, given that there is no unique mapping from the real price of oil to the underlying structural shocks. The nature of the structural shocks determining the path of the future real price of oil, however, is what determines the implied future path of the other model variables. Thus, without specifying the precise sequence of structural shocks, the scenario analysis would be ill-defined.

## 4.5 Simulating Counterfactual Outcomes

The construction of historical decompositions and of conditional forecasts is closely related to the construction of counterfactuals, as discussed in Kilian (2017). A counterfactual refers to a simulation of the path of the VAR model variables under a different sequence of structural shocks than observed in the actual data. The use of counterfactuals is best illustrated by an empirical example.

There has been a remarkable surge in the production of unconventional crude oil in the United States since November 2008. This surge has been tied to the rapid expansion of the production of shale oil (see Kilian 2016). The U.S. shale oil boom has important ramifications for the global market for crude oil. It represents a shift in the global supply curve along the global demand curve, so, all else equal, one would expect the shale oil boom to have lowered the global price of crude oil, as measured by the Brent price of crude oil. What is not obvious is how much higher the Brent price would have been in the absence of the shale oil boom. Answering this question requires the construction of a counterfactual based on a structural dynamic model of the global oil market.

In short, the question of interest is how different the global price of crude oil would have been if all producers other than the United States had maintained their observed oil production levels but the U.S. shale oil boom had never happened. Kilian (2017) provides an answer to this question based on a thought experiment in which increased U.S. shale oil production is entirely due to exogenous oil supply shocks rather than being caused by oil demand shocks. This assumption is consistent with the view that in the absence of technological innovation, the shale oil boom would not have taken place, but it makes no allowance for any additional effects working through oil price expectations.

The analysis involves three steps. First, we construct the counterfactual level of global oil production. Second, we specify a structural VAR model of the global oil market and estimate this model on the full sample. Given the change in the quantity of oil produced in equilibrium under the counterfactual, it is

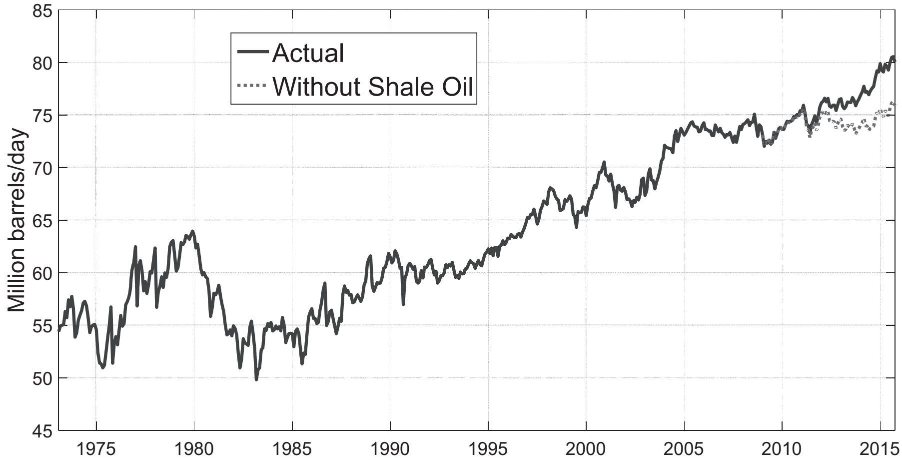


**Figure 4.7.** Actual level of global oil production and counterfactual level in the absence of the U.S. shale oil boom.
*Source:* Kilian (2017).

straightforward to infer from this model how much the supply curve must shift each month, conditional on past data, in order to produce the counterfactual equilibrium quantity of global oil production. Third, we replace the original sequence of flow supply shocks by the resulting counterfactual sequence of flow supply shocks into the structural VAR model and simulate the evolution of the global price of crude oil under this new sequence, while leaving unchanged all other structural shocks in the model.

The construction of the counterfactual level of global oil production in step 1 relies on direct measures of the volume of shale oil production available from the U.S. Energy Information Administration. Figure 4.7 shows the actual evolution of global oil production as well as the counterfactual path obtained by subtracting the barrels of shale oil produced since November 2008 from the actual data.

In step 2, Kilian (2017) employs an updated version of the global oil market model of Kilian and Lee (2014) that already served as an empirical example in Sections 4.3 and 4.4. Let $y_t = (y_{1t}, y_{2t}, y_{3t,} y_{4t})'$ denote the vector of variables in this model, where $y_{1t}$ stands for the growth in world oil production and $y_{3t}$ for the log of the real price of crude oil, in particular. Consider the structural VAR model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_{24} y_{t-24} + w_t,$$

where the $4 \times 1$ vector $y_t$ is assumed to be stationary and the deterministic regressors have been suppressed for expository purposes. The dimension of

$B_i$, $i = 0, \ldots, 24$, is $4 \times 4$. The $4 \times 1$ vector $w_t = (w_{1t}, w_{2t}, w_{3t,} w_{4t})'$, where $w_{1t}$ denotes a shock to the global flow supply of crude oil, is assumed to be zero mean white noise with a diagonal $4 \times 4$ covariance matrix $\Sigma_w$ that is of full rank and that, without loss of generality, can be normalized to equal the identity matrix. This structural model can be expressed in reduced form as

$$y_t = A_1 y_{t-1} + \cdots + A_{24} y_{t-24} + u_t,$$

where $A_i = B_0^{-1} B_i,\ i = 1, \ldots, 24$, and $u_t = B_0^{-1} w_t$ with variance-covariance matrix $\Sigma_u = B_0^{-1} \left(B_0^{-1}\right)'$. Let the $4 \times 4$ matrix $\partial y_{t+i}/\partial w_t' = \Theta_i$ denote the responses of the model variables to each of the structural shocks at horizon $i = 0, 1, \ldots, H$. The matrix $\Theta_i = \left[\theta_{jk,i}\right]$ consists of the elements $\theta_{jk,i} = \partial y_{j,t+i}/\partial w_{kt}$ that denote the response of variable $j$ to structural shock $k$ at horizon $i$. These responses may be computed as $\Theta_i = J\mathbf{A}^i J' B_0^{-1}$, where $J = \left[I_4, 0_{4\times 4(24-1)}\right]$ and $\mathbf{A}$ denotes the matrix of slope parameters obtained by expressing the VAR(24) model in its VAR(1) companion format.

In constructing the counterfactual in question, we make use of the fact that, after removing the deterministic terms,

$$y_t \approx \sum_{i=0}^{t-1} \Theta_i w_{t-i},$$

as already noted in the section on historical decompositions. As a result, the fitted value for the log real price of oil can be written as

$$\widehat{y}_{3t} = \sum_{j=1}^{4} \widehat{y}_{3t}^{(j)},$$

where $\widehat{y}_{3t}^{(j)}$ denotes the cumulative contribution since the beginning of the sample of structural shock $j$ to the third model variable at time $t$, defined as

$$\widehat{y}_{3t}^{(1)} \equiv \sum_{i=0}^{t-1} \theta_{31,i} w_{1,t-i},$$

$$\widehat{y}_{3t}^{(2)} \equiv \sum_{i=0}^{t-1} \theta_{32,i} w_{2,t-i},$$

$$\widehat{y}_{3t}^{(3)} \equiv \sum_{i=0}^{t-1} \theta_{33,i} w_{3,t-i},$$

$$\widehat{y}_{3t}^{(4)} \equiv \sum_{i=0}^{t-1} \theta_{34,i} w_{4,t-i}.$$

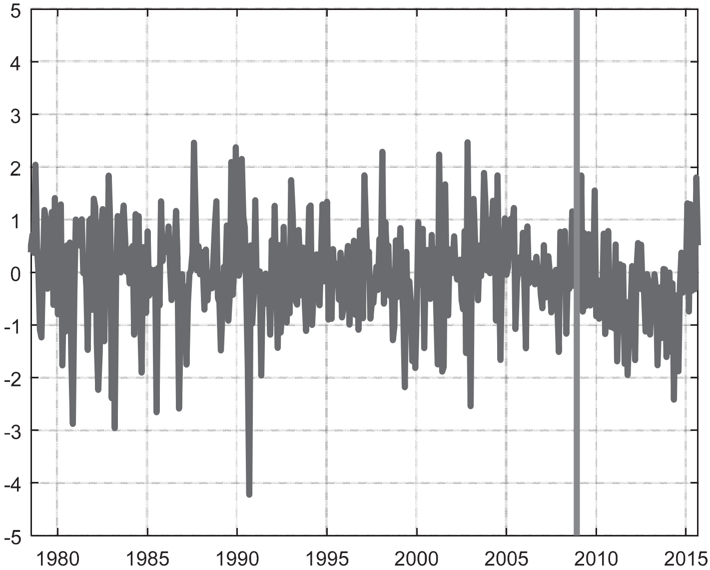


**Figure 4.8.** Counterfactual sequence of flow supply shocks in the absence of the U.S. shale oil boom.
*Notes:* The vertical line marks the beginning of the shale oil boom in November 2008.
*Source:* Kilian (2017).

In practice, $\theta_{jk,i}$ and $w_t$ are replaced by consistent estimates. An analogous decomposition also exists for the fitted value $\widehat{y}_{1t}$ of the growth rate of global oil production. These decompositions play an important role in constructing the counterfactual.

Having recovered estimates of $B_0^{-1}$, $\theta_{jk,i}$ and $w_t = B_0 u_t$, $t = 1, \ldots, T$, from the full-sample estimates of the Kilian and Lee (2014) model, as discussed in Section 4.3, we retain the realizations of the flow supply shock, $w_{1t}$, up to November 2008, but replace the estimates of $w_{1t}$ for the remainder of the sample by counterfactual values chosen to ensure that the path of global oil production corresponds to the counterfactual path shown in Figure 4.7. Given the demeaned growth rate of world oil production implied by the counterfactual, these values may be easily computed by an iterative procedure. Having computed what the expected value of $y_{1t}$, given $\widehat{y}_{1t}^{(j)}$, $j = 2, 3, 4$, would have been next month if this month's supply shock had been zero, the difference between that model prediction and the target level of oil production growth, scaled by the impact response of world oil production to a flow supply shock, will be the magnitude of the flow supply shock, $w_{1t}$, required to reach the target. Figure 4.8 shows the sequence of flow supply shocks that would have to be imposed in the oil market VAR model to make global oil production follow the counterfactual path of world oil production in Figure 4.7. By construction, this shock sequence is identical to the observed structural residuals, $\widehat{w}_{1t}$, until November 2008.

An important concern is whether the flow supply shocks required to implement the prespecified path of global oil production after November 2008 are too large by historical standards or too predictable to maintain the assumption of a time-invariant structural VAR model. Figure 4.8 illustrates that the counterfactual shock sequence involves no flow supply shocks that are unusually large by historical standards, addressing the first concern. Nor is there reason to believe that economic agents would have been able to predict these counterfactual flow supply shocks. This concern would be warranted if there were an unusually large number of consecutive flow supply shocks of the same sign (referred to as a "run") relative to past data. The longest run of flow supply shocks observed under the counterfactual after December 2008 is fourteen months. Although the longest run in the flow supply shock sequence prior to December 2008 lasted only nine months, a run of fourteen months is not entirely unprecedented. Runs of similar length can be found in other structural shock sequences in the same model, suggesting that this counterfactual is not unreasonable.

To answer the question of how much the global price of oil would have changed in the absence of the shale oil boom, in step 3 we recompute $\widehat{y}_{3t}^{(1)}$ under the counterfactual sequence of flow supply shocks while retaining the other three structural shock sequences, as originally estimated, in computing $\widehat{y}_{3t}^{(2)}$, $\widehat{y}_{3t}^{(3)}$, and $\widehat{y}_{3t}^{(4)}$. Comparing the implied sequence of $\widehat{y}_{3t}$ to the demaned value of $y_{3t}$ provides a measure of the cumulative impact of the shale oil boom on the log real price of oil.

In order to map the results for the log real price of oil (defined as the real U.S. refiners' acquisition cost for oil imports in the Kilian and Lee model) to the nominal Brent price, we proceed as follows. Given the implied sequence of $\widehat{y}_{3t}$, the counterfactual Brent price is constructed from the real price of Brent crude oil, scaled up by the percent deviation between the counterfactual real price of oil and the actual real price of oil in the VAR model. The counterfactual real Brent price is then converted back to dollars using the U.S. consumer price index. Figure 4.9 shows how much higher the nominal Brent price of crude oil would have been in the absence of the shale oil boom.

The gap between the actual and the counterfactual path of the Brent price measures the cumulative effect of the shale oil boom on the Brent price. This effect became important starting in 2011 and reached nearly 10 dollars per barrel of crude oil in early 2014, before declining toward the end of the sample.

This example illustrates how counterfactuals may be used to answer policy-relevant questions based on structural VAR models. The same approach could be used, in principle, to simulate the performance of the economy under a

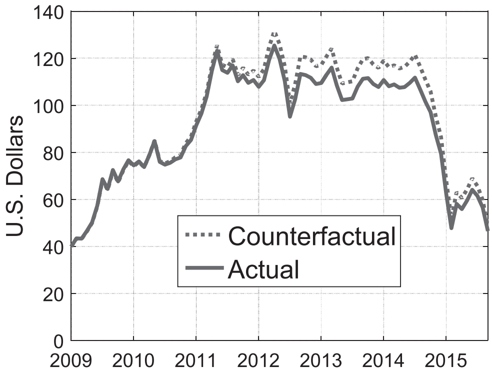


**Figure 4.9.** Evolution of the nominal Brent price of crude oil with and without shale oil.
*Source:* Kilian (2017).

counterfactual path of the interest rate for a given policy reaction function.[2] The construction of policy counterfactuals becomes more complicated when the counterfactual involves changing the policy rule rather than merely the path of the interest rate conditional on the estimated policy reaction function. This situation is considered in the next subsection.

## 4.6 Policy Counterfactuals

An important question in structural VAR analysis dating back to Bernanke, Gertler, and Watson (1997) and Sims and Zha (2006a), among others, has been how to construct policy counterfactuals. For example, Bernanke et al. were interested in assessing the effects of the systematic response of monetary policy to exogenous oil price shocks. For this purpose they augmented the monetary policy reaction function in an otherwise standard monetary policy VAR model to allow the policymaker to respond directly to oil price shocks. Their premise was that this systematic policy response was responsible in part for the subsequent recessions. The policy counterfactual Bernanke et al. proposed involved the central bank holding the interest rate constant following the oil price innovation. The question of interest was how recessionary the consequences of oil price shocks would have been, had the central bank followed a constant interest rate rule.

A different policy counterfactual for the same class of models was proposed by Kilian and Lewis (2011). They observed that – if we are interested in the

[2] Note that one could also construct a forecast conditioning on the path of the interest rate in much the same way, as discussed in Section 4.4.2, except that in the latter case future values of the remaining structural shocks would be set to zero in expectation rather than being held fixed at their actual realizations.

question of how much of a difference the direct response of the central bank to oil price shocks makes – the relevant benchmark should be a conventional model of monetary policy in which the central bank reacts to fluctuations in other macroeconomic variables such as inflation and real output with only the direct response to oil price shocks shut down.

In what follows, we contrast the construction of these two counterfactuals, referred to as the BGW and KL counterfactual for convenience. For illustrative purposes, we consider a VAR model that includes only real non-oil commodity prices, the real price of oil, U.S. real output, U.S. inflation, and the federal funds rate. Under the KL counterfactual of only shutting down the direct response to the real price of oil, the objective is to construct a sequence of hypothetical shocks to the federal funds rate that offsets the contemporaneous and lagged effects of including the real price of oil in the policy reaction function. Such a shock sequence may be recovered from the estimated VAR, allowing us to compare the actual evolution of the VAR variables during a given episode with that under the counterfactual policy shock sequence.

In contrast, for the original BGW counterfactual, which Bernanke et al. refer to as the Sims-Zha counterfactual based on a proposal dating back to 1996 and published in Sims and Zha (2006a), one simply constructs a hypothetical path of the policy shock that offsets all endogenous dynamics in the federal funds rate such that the federal funds rate remains unchanged over time. For a more detailed description of the construction of these counterfactuals, see Kilian and Lewis (2011). For an alternative description of the BGW counterfactual, see Hamilton and Herrera (2004).

Figure 4.10 provides an illustration of the two counterfactuals based on the VAR model of monetary policy estimated in Kilian and Lewis (2011) on monthly data for 1967m5–1987m7. It compares the evolution of real output (measured by the Chicago Fed National Activity Index or CFNAI), of consumer price inflation, and of the federal funds rate in response to a one-time unexpected increase in the real price of oil under each of the two counterfactuals. It also shows the response of these observables under the policy regime maintained in the unrestricted VAR model. Figure 4.10 shows that shutting down the direct response to oil price shocks has virtually no effect on inflation and little effect on real output. The Federal Reserve still would have raised interest rates by roughly the same number of basis points in response to an exogenous positive oil price shock, but the bulk of that response would have occurred three months later. Holding the interest rate constant, in contrast, as proposed by Bernanke et al., again would have had virtually no effect on inflation and real output would have been only slightly higher than observed.

An obvious concern in constructing policy counterfactuals from structural VAR models is that such counterfactuals may run afoul of the Lucas critique (see Lucas 1976). The concern is that agents may change their behavior in

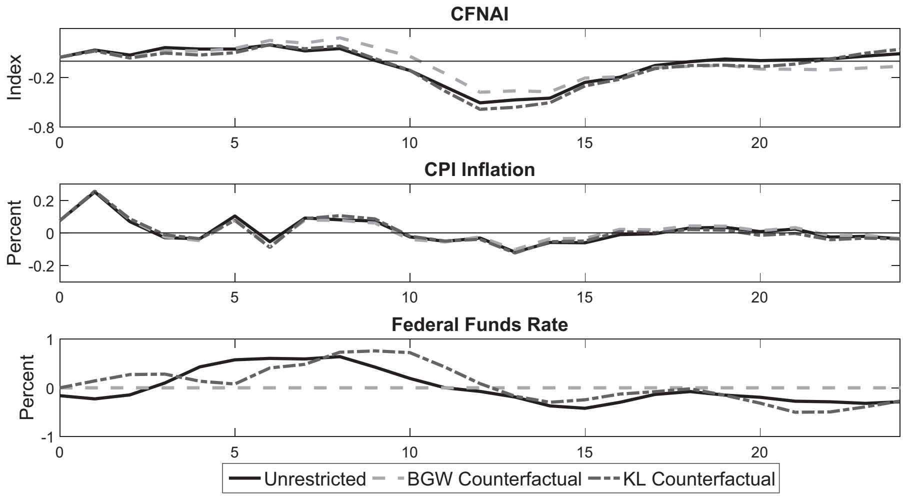


**Figure 4.10.** Counterfactual paths of key observables under alternative policy counterfactuals.
*Source:* Kilian and Lewis (2011).

anticipation of shifts in policy, causing the structure of the VAR model to change over time, thus invalidating the analysis. A reasonable presumption is that a policy shock sequence outside of the range of historical experience is likely to be recognized by agents and, if implemented, would cause changes in the structure of the underlying model. Leeper and Zha (2003) make the case that modest policy interventions, in contrast, do not significantly shift agents' beliefs about the policy regime and hence are not subject to the Lucas critique.

Hamilton and Herrera (2004) discuss diagnostics for evaluating whether the shock sequences required to implement counterfactual paths for policy instruments in VAR models are subject to the Lucas critique. They focus on both the magnitude of the policy surprises required to implement the counterfactual and the extent to which agents would have been able to anticipate these surprises. Their conclusion is that the constant interest rate counterfactual of Bernanke et al. involves large policy shocks outside of the range of historical experience and requires agents to be fooled by repeated policy surprises in the same direction for an extended period. Applying these criteria to the counterfactual proposed by Kilian and Lewis (2011) and illustrated in Figure 4.10 does not reveal the same problem. This point is illustrated in Figure 4.11 which plots the sequences of policy shocks required to implement this counterfactual, given the monetary policy regime in the unrestricted structural VAR model. Not only are the hypothesized interest rate shocks well within the range of historical

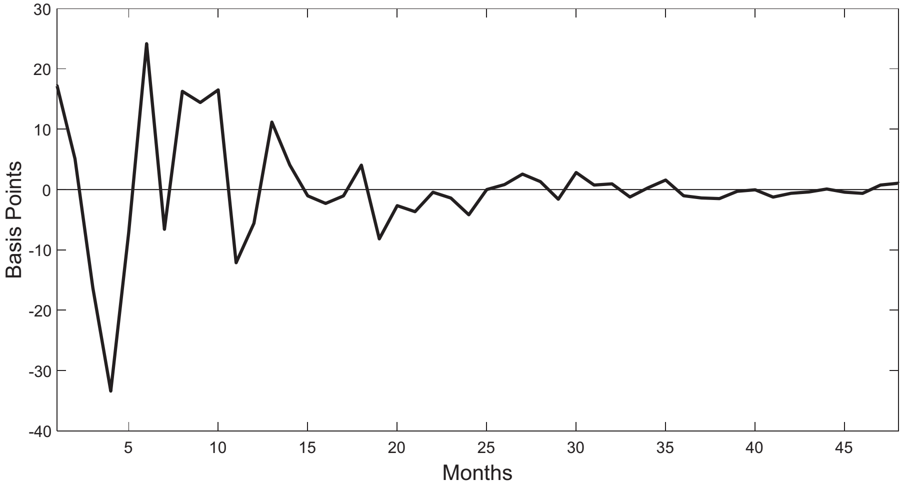


**Figure 4.11.** Sequence of policy shocks required to implement the KL policy counterfactual.
*Source:* Adapted from Kilian and Lewis (2011).

policy shocks but there is no evidence of serial correlation in the policy shock sequence that would allow agents to predict and anticipate the counterfactual path of monetary policy.

These examples illustrate that there are obvious limits to the types of policy counterfactuals that may be entertained based on a structural VAR model. Nevertheless, when used with care, policy counterfactuals may provide valuable insights.