17 Nonfundamental Shocks

17.1 Introduction

In Chapter 2 we motivated VAR models as approximations to a DGP of the form

$$y_t = \Phi(L)u_t,$$

where u_t is the one-step ahead prediction error based on information $\{y_{t-1}, y_{t-2}, \dots\}$. Structural VAR analysis is based on the premise that the structural shocks can be obtained by transforming the prediction errors u_t. The MA representation based on the VAR prediction errors is known as a fundamental representation. It is conceivable, however, that the shocks constituting the MA representation of the DGP do not coincide with the prediction errors of a given VAR model. For example, the DGP may be

$$y_t = \Phi^*(L)u_t^*,$$

where u_t^* is not the prediction error based on $\{y_{t-1}, y_{t-2}, \dots\}$. In this case, there are two MA representations of the same time series, one of which involves the fundamental shock u_t, whereas the other involves the nonfundamental shock u_t^*. These two MA representations have the same first and second moments, but their MA operators $\Phi(L)$ and $\Phi^*(L)$ differ. If the coefficients of $\Phi^*(L)$ reflect the dynamic multipliers of shocks, the econometrician's effort to recover the true structural impulse responses from the prediction error u_t will be doomed. This chapter reviews how this problem of nonfundamental shocks u_t^* may arise, how it may be detected, and how it may be resolved or at least ameliorated.[1]

Nonfundamental shocks have been the subject of a growing number of studies on structural macroeconometric modeling. Early theoretical work on nonfundamentalness dates back to Hansen and Sargent (1980, 1991). An application illustrating the potential importance of nonfundamental

[1] This chapter is based in part on Lütkepohl (2014).

representations is provided in Lippi and Reichlin (1993). Nonfundamental representations of stochastic processes are also discussed in Blanchard and Quah (1993) and Lippi and Reichlin (1994). More recent work on nonfundamental representations and shocks includes Giannone and Reichlin (2006), Fernández-Villaverde and Rubio-Ramírez (2006), Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), Forni, Giannone, Lippi, and Reichlin (2009), and Canova and Hamidi Sahneh (2016). A review of that literature is provided by Alessi, Barigozzi, and Capasso (2011).

The main argument in favor of nonfundamental shocks being important in economic analysis is that the econometrician may not have all the information that economic agents have. In particular, expectations of households, firms, and policymakers may not be based on past information from the variables in the econometric model alone, but may also reflect additional information from outside this model. Hence, the prediction errors governing the DGP cannot be equated with the reduced-form innovations of the VAR model set up by the econometrician. As a result, the shocks of economic interest may be nonfundamental. For example, Leeper, Walker, and Yang (2013) consider a model of fiscal policy shocks in which agents are assumed to know about tax changes in advance. This feature gives rise to nonfundamental tax shocks that undermine conventional estimates of the tax multiplier.

In this chapter, we provide a review of this literature and make the case that some practical responses to possible nonfundamentalness may be more appealing than others. Two common responses to the nonfundamentalness problem are either to allow for noninvertible MA components or to view nonfundamentalness as an omitted-variables problem. In the latter case, additional variables must be added to the model to align the information set used by the econometrician with the agents' information set. For example, Ramey and Shapiro (1998), Chung and Leeper (2007), Romer and Romer (2010), and Ramey (2011) use this approach in the context of fiscal policy analysis (see Chapter 7). We make the case that the omitted variables problem encompasses the nonfundamentalness problem. Solving it will also solve the nonfundamentalness problem, whereas solving the nonfundamentalness problem without adding further variables will not solve all of the problems caused by omitted variables. We emphasize that using factor models or Bayesian techniques to allow for larger information sets, as has recently been proposed, may not suffice to resolve the omitted variables problem. In some cases an alternative solution is to augment the model by a judiciously chosen variable that encompasses the missing information (see, e.g., Leeper, Walker, and Yang 2013; Kilian and Murphy 2014).

In Section 17.2 we formally define fundamental and nonfundamental representations of stochastic processes and the corresponding shocks. In Section 17.3 we review the literature on fundamental and nonfundamental representations. We conclude in Section 17.4.

17.2 Fundamental and Nonfundamental Moving Average Representations

In this section, we focus on a purely nondeterministic process y_t because deterministic terms are not needed for the subsequent arguments. Moreover, we exclude processes with MA roots on the unit circle and cointegrated processes as well as other types of nonstationarities. Some of these features are important in practice, but they do not affect the main arguments in this chapter. Consider a K-dimensional stationary stochastic process y_t with MA representation

$$y_t = \Phi(L)u_t, \quad t \in \mathbb{Z}, \tag{17.2.1}$$

where $u_t \sim (0, \Sigma_u)$ is a zero-mean M-dimensional white noise process with time-invariant nonsingular covariance matrix Σ_u and uncorrelated u_t and u_s for $t \neq s$. By allowing the dimension of the white noise process to be different from the dimension of y_t we generalize the framework of previous chapters. The operator

$$\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i \tag{17.2.2}$$

is a possibly infinite order $K \times M$ matrix power series in the lag operator L, with absolutely summable coefficient matrices Φ_i. This setup covers cases in which y_t admits VAR or VARMA representations.

What is Fundamentalness? The MA representation is fundamental if u_t is the one-step ahead prediction error associated with the optimal linear prediction of y_t based on lagged y_t. Our formal definition of fundamentalness and nonfundamentalness follows from Rozanov (1967) and Alessi, Barigozzi, and Capasso (2011). Suppose that all random variables of interest belong to the Hilbert space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ based on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as in Alessi, Barigozzi, and Capasso (2011), where Ω is the sample space, $\mathcal{F}$ is the σ-algebra describing all possible events, and $\mathbb{P}$ is the probability measure. The symbol $\mathcal{H}_t^y$ denotes the subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by $\{y_t, y_{t-1}, \dots\}$.

Definition. The process u_t is *y_t-fundamental* if $\mathcal{H}_t^y = \mathcal{H}_t^u$ for all $t \in \mathbb{Z}$. The process u_t is *y_t-nonfundamental* if $\mathcal{H}_t^y \subset \mathcal{H}_t^u$ and $\mathcal{H}_t^y \neq \mathcal{H}_t^u$ for any $t \in \mathbb{Z}$.

In other words, u_t is y_t-fundamental if it can be written as a function of $y_t, y_{t-1}, \dots$; otherwise it is y_t-nonfundamental. Using results of Rozanov (1967), Forni, Giannone, Lippi, and Reichlin (2009) observe that u_t is y_t-fundamental if, for $M \leq K$, rank$(\Phi(z)) = M$ for all z in the complex unit circle. Thus, for $M = K$, if all roots of $\det(\Phi(z))$ are outside the complex unit circle, expression (17.2.1) is a fundamental representation that can be inverted to obtain a VAR representation of possibly infinite order. The u_t in this case is the one-step ahead prediction error associated with the optimal linear prediction of y_t based on $\mathcal{H}_{t-1}^y$.

The Relationship Between the Fundamental and Nonfundamental Representations. If (17.2.1) is a fundamental MA representation, then there are also nonfundamental MA representations with the same first and second moments. For example, if u_t is a scalar zero-mean white noise process with variance σ_u^2, the scalar MA(1) process $y_t = u_t + \phi u_{t-1}$ is y_t-fundamental if $|\phi| < 1$. There is an equivalent nonfundamental representation $y_t = u_t^* + \frac{1}{\phi} u_{t-1}^*$, where u_t^* is zero-mean white noise with variance $\sigma_{u^*}^2 = \phi^2 \sigma_u^2$. It is easy to verify that this representation implies the same autocovariance structure for y_t. Clearly, the MA polynomial $1 + \frac{1}{\phi} L$ has a root $-\phi$ inside the unit circle. Hence, u_t^* cannot be obtained from y_t and its lags by inverting the MA polynomial.

For more general K-dimensional fundamental processes y_t, we obtain

$$y_t = \Phi(L)\mathbb{B}(L)\mathbb{B}(L)^{-1}u_t = \Phi^*(L)u_t^*, \tag{17.2.3}$$

where $u_t^* = \mathbb{B}(L)^{-1}u_t$. The MA operator $\Phi^*(z)$ has roots inside the complex unit circle if $\mathbb{B}(L)$ is chosen to be a so-called Blaschke matrix (see Lippi and Reichlin 1994). A matrix operator $\mathbb{B}(z)$ is a Blaschke matrix if it has no poles in and on the complex unit circle and $\mathbb{B}(z)^{-1} = \mathbb{B}^*(\frac{1}{z})$, where the asterisk denotes the conjugate transpose. For example,

$$\begin{bmatrix} \dfrac{z-a}{1-\bar{a}z} & 0 \\ 0 & I_{K-1} \end{bmatrix}$$

with $|a| < 1$ is a Blaschke matrix. Here $\bar{a}$ denotes the complex conjugate of a. An important property of Blaschke matrices is that $u_t^* = \mathbb{B}(L)^{-1}u_t$ is white noise if u_t is white noise.

In the nonfundamental representation (17.2.3) the white noise process u_t^* is not the optimal linear prediction error of y_t, given lagged y_t. Note, however, that nonfundamental MA representations are equivalent to the fundamental MA representation in that they have the same second-order properties and, hence, represent the same stochastic process y_t under Gaussianity. In a non-Gaussian setting, in contrast, it is possible to discriminate between fundamental and nonfundamental representations based on higher moments, as shown in Huang and Pawitan (2000), Chan and Ho (2004), and Gourieroux and Monfort (2014).

Nonfundamental representations are also related to noncausal VAR models. Noncausal VAR models include regressors y_{t+i} in the equation for y_t. Such models have nonfundamental shocks by construction (see, e.g., Brockwell and Davis (1987, Chapters 3 and 11) and Lanne and Saikkonen (2013) for discussions of noncausal processes). Lanne and Saikkonen (2013) propose statistical procedures for discriminating between causal and noncausal processes in practice.[2]

17.3 Fundamental versus Nonfundamental Representations

MA representations are directly linked to impulse response analysis, which is the main tool used in economic analysis. The impulse responses based on nonfundamental MA representations may be quite different from those based on the fundamental MA representation. This fact explains why the question of fundamentalness is considered important in macroeconometrics. It should also be noted that the nonfundamental MA representation may be quite different from the generalized impulse responses proposed by Koop, Pesaran, and Potter (1996) (see Chapter 18). The latter impulse responses are based on conditional expectations conditioning on past observations, whereas under nonfundamentalness such conditional expectations would be nonlinear functions that do not correspond to the coefficients of the MA lag polynomial.

17.3.1 Nonfundamental Shocks in Economic Models

The main argument in favor of nonfundamental representations is that the information set of economic agents may be larger than the information set available to the econometrician. In this case, agents' expectations differ from the conditional expectations obtained from the econometrician's model for y_t. Since the shocks underlying the fundamental MA representation are the econometrician's prediction errors, they will differ from the shocks of interest that would be obtained on the basis of the agents' larger information set.

Example: Fiscal Foresight. Leeper, Walker, and Yang (2013) consider a simple model for assessing the effects of fiscal policy shocks. The model implies that without fiscal foresight the deviation of the capital stock from its steady state value in quarter t, k_t, in equilibrium depends only on an iid technology shock, $w_t^{\text{technology}}$,

$$k_t = \rho k_{t-1} + w_t^{\text{technology}},$$

with capital accumulation being unaffected by iid shocks to lump-sum taxes w_t^{tax}. In contrast, with some degree of tax foresight, even serially uncorrelated tax hikes reduce capital accumulation. For example, if agents know the tax shock two quarters in advance,

$$k_t = \rho k_{t-1} + w_t^{\text{technology}} - \kappa(w_{t-1}^{\text{tax}} + \theta w_t^{\text{tax}}),$$

[2] The notion of causality in this literature is not the same as the standard notion of economic causality discussed in Chapter 7.

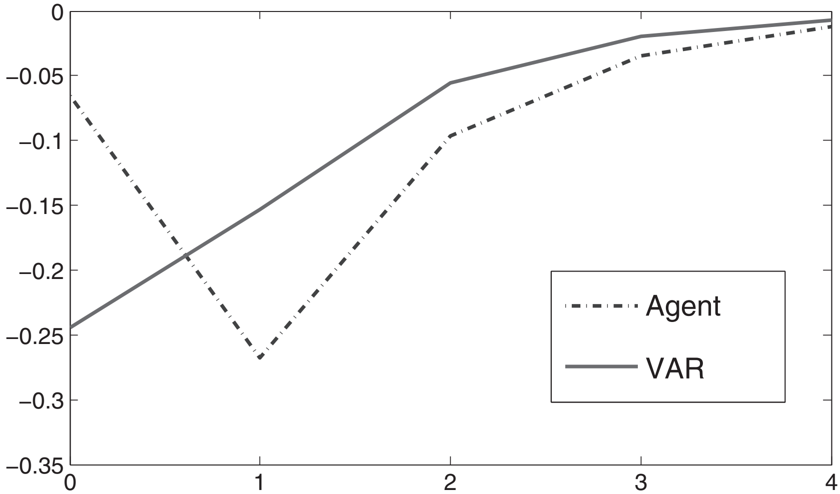

Figure 17.1. Response of the capital stock to an iid tax shock.
Source: Adapted from Leeper, Walker, and Yang (2013) with the help of MATLAB code kindly made available by Eric Leeper.

where $\kappa = (1-\theta)(\tau/(1-\tau))$, τ is the tax rate in steady state and $\theta < 1$ is a discount factor.

Leeper, Walker, and Yang (2013) prove that – abstracting from the technology shock – what the econometrician recovers as the tax innovation at time t by fitting an autoregression based on current and past observations for k_t is actually a discounted sum of the tax shocks observed by agents at date t or earlier. Thus, the tax shocks w_t^{tax} are not fundamental. Ignoring this fact results in distorted impulse responses. Figure 17.1 shows the difference between the response of the capital stock estimated by the econometrician and the true response implied by the model with fiscal foresight. Intuitively, this discrepancy arises because the econometrician attributes all of the dynamics associated with the anticipated component of the tax rate to the unanticipated component. The largest response to the tax shock occurs instantaneously in the econometrician's model. In contrast, in the model with tax foresight, the largest response of capital is seen to be one period after the arrival of the shock.

This simple example illustrates that, more generally, VAR models must account for forward-looking behavior to avoid spurious conclusions about structural impulse responses and forecast error variance decompositions. The discussion of fiscal policy in Leeper, Walker, and Yang (2013) is one prominent example of a situation in which economic foresight cannot be ignored. Economic foresight also plays an important role in modeling asset prices and commodity prices, for example. This realization has prompted concerns about the reliability of structural VAR analysis in many applications.

The central problem is that inverting the VAR lag operator by construction yields a fundamental MA representation, regardless of whether the DGP is fundamental or not. Two general approaches have been proposed to cope with the nonfundamentalness problem. One is to allow for MA representations in which the determinant has roots inside the complex unit circle when evaluating the VAR model, thereby explicitly accounting for the possibility of nonfundamental shocks. This approach requires suitable transformations of the model. The other approach is to broaden the information set by increasing the number of variables and, hence, the information included in the VAR model. The latter solution has been favoured by advocates of factor-augmented VAR (FAVAR) models in particular. As discussed in Chapter 16, FAVAR models can accommodate much larger sets of variables than conventional unrestricted VAR models, as can VAR models estimated by suitable Bayesian methods.

In Sections 17.3.2 and 17.3.3 we discuss these proposals in general, followed by specific comments in Section 17.3.4 on the use of FAVAR and large-scale BVAR models as a means of dealing with nonfundamentalness. In Section 17.3.5 we discuss other options for dealing with nonfundamental shocks that may be available in specific circumstances.

17.3.2 Nonfundamentalness Due to MA Roots in the Unit Circle

A fundamental MA representation can be turned into a nonfundamental one by applying the Blaschke matrices mentioned in Section 17.2 and, conversely, nonfundamental representations can be transformed into fundamental ones with equivalent second-order properties. In fact, if the process has a nonfundamental MA representation of the form (17.2.3) such that $\det(\Phi^*(z))$ has no roots on the unit circle, there always exists an equivalent fundamental MA representation. Thus, the question is not whether the true process is fundamental or nonfundamental (e.g., Alessi, Barigozzi, and Capasso 2011, p. 11). The process is always fundamental in the sense that it admits a fundamental representation as long as the process is Gaussian or only the first and second moments are of interest. Nevertheless, it may be the case that the shocks of interest for impulse response analysis are nonfundamental.

In a Gaussian setting, a VAR with fundamental errors can be fitted and used as the basis for determining the nonfundamental representation of interest as discussed in Section 17.2. For example, suppose that a univariate variable y_t admits an MA(1) representation,

$$y_t = u_t + \phi u_{t-1}, \tag{17.3.1}$$

where y_t-fundamentalness requires that $|\phi| < 1$. As mentioned in Section 17.2, an equivalent nonfundamental representation is $y_t = u_t^* + \frac{1}{\phi} u_{t-1}^*$. For the MA(1) process in (17.3.1) this is actually the only basic nonfundamental representation in the sense of Lippi and Reichlin (1994). Clearly, $|\frac{1}{\phi}| > 1$ and, hence, the response of y_t to a nonfundamental shock after one period can be larger than 1, whereas the corresponding response to a fundamental shock is smaller than 1.

Whether it is useful or not to solve the nonfundamentalness problem by allowing for MA roots in the unit circle depends on whether the economic theory underlying the analysis requires nonfundamental representations. Obviously, allowing for nonfundamental representations in addition to fundamental ones potentially makes identification of the structural shocks of interest more difficult because it opens up more possibilities. If only fundamental shocks are considered, the structural shocks of interest are obtained by linearly transforming the residuals of the reduced-form VAR. If, however, in addition nonfundamental shocks are potentially of interest, all nonfundamental representations obtained by Blaschke matrix transformations plus all admissible instantaneous linear transformations must be considered when selecting a model representation, which may be a problem if economic theory does not deliver a full set of identifying restrictions. This is not a problem of the VAR or VARMA processes being the appropriate stochastic model for the DGP, but simply a problem of choosing the right representation for economic analysis.

It is worth mentioning that statistical tools can also help in determining whether a nonfundamental representation is required. As pointed out by Huang and Pawitan (2000), Davis and Song (2010), Lanne and Saikkonen (2013), and Gourieroux and Monfort (2014), nonfundamental and fundamental representations are equivalent for Gaussian processes whose distribution is determined completely by their first and second moments. For non-Gaussian processes fundamental and nonfundamental representations are not equivalent in that they give rise to processes that can be distinguished by their higher order moments. Hence, statistical procedures can help to discriminate between fundamental and nonfundamental representations. In particular, they may indicate that a nonfundamental representation is more suitable. Of course, they may also lead to the conclusion that a nonfundamental representation is not supported by the data.

17.3.3 Nonfundamentalness Due to Omitted Variables

More commonly, nonfundamental shocks are associated with an omitted-variables problem. If the econometrician considers a set of variables y_t, which does not contain all variables of importance to the economic agents, the VAR may still be a perfectly valid representation of the DGP of y_t, but due to omitted variables may not properly capture the structural impulse responses of economic interest. In other words, the system of interest for an economist may be $z_t = (y_t', x_t')'$, which contains the vector y_t of original VAR variables as well as a vector x_t of additional variables that are omitted from the original VAR specification. Suppose

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Theta_{yy}(L) & \Theta_{yx}(L) \\ \Theta_{xy}(L) & \Theta_{xx}(L) \end{bmatrix} \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix} = \Theta(L)u_{zt} \tag{17.3.2}$$

is a fundamental MA representation of z_t. Here the partitioning of the matrix polynomial in the lag operator $\Theta(L)$ and the residual term u_{zt} corresponds to the partitioning of z_t. Clearly, $u_{zt} = (u'_{yt}, u'_{xt})'$ generally cannot be recovered from past and present y_t, i.e., $\mathcal{H}_t^y$ is a strict subset of $\mathcal{H}_t^{u_z}$ in general and, hence, u_{zt} is not y_t-fundamental.

If a u_{yt} shock hits the system (17.3.2), the marginal responses of y_t are given by the appropriate elements of $\Theta_{yy}(L)$, which can be quite different from the impulse responses obtained from $\Phi(L)$, which denotes the MA operator in (17.2.1), for the simple reason that different shocks are considered. Of course, this insight is not new (see, e.g., Lütkepohl 1991, section 2.3.2; Giannone and Reichlin 2006). Note, however, that both MA representations (17.2.1) and (17.3.2) are fundamental, given their respective information sets: u_t in (17.2.1) is y_t-fundamental and u_{zt} in (17.3.2) (and hence, u_{yt}) is z_t-fundamental. Thus, the relevant MA operators have all roots outside the complex unit circle, which is a very different solution to the nonfundamentalness problem than the one discussed earlier.

Testing for Omitted Variables. Giannone and Reichlin (2006) make the point that the nonfundamentalness problem arises if any of the additional variables x_t is Granger causal for y_t (see Chapter 2). As is well known, if the MA representation is normalized such that $\Theta(0)$ is the identity matrix, x_t is Granger noncausal for y_t if and only if $\Theta_{yx}(L) = 0$ (see, e.g., Lütkepohl 1991, section 2.3.1, proposition 2.2). Thus, it is obvious that u_{yt} is y_t-fundamental, if $\Theta_{yx}(L) = 0$ and, hence, the MA operator in model (17.3.1) is lower block-triangular. In other words, if x_t is not Granger causal for y_t, u_{yt} can be recovered from present and past y_t. Conversely, if x_t is Granger causal for y_t, u_{yt} is not y_t-fundamental. A test along these lines has been conducted, for example, in Kilian and Murphy (2014) to address the question of whether the oil futures spread contains additional forward-looking information not already captured by the variables of a model of the physical market for crude oil. Canova and Hamidi Sahneh (2016) observe that Granger-noncausality tests may provide spurious evidence of nonfundamentalness under certain forms of model misspecification. They provide an alternative and more robust test of nonfundamentalness.

Why We Need More Information. To see that increasing the information set directly may be a better way of thinking about solving the fundamentalness problem than just allowing for MA roots in the unit circle, consider a simple illustrative example. Suppose y_t and x_t are both scalar variables that are jointly generated by the bivariate MA(1)

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix} + \begin{bmatrix} \theta_{yy} & \theta_{yx} \\ \theta_{xy} & \theta_{xx} \end{bmatrix} \begin{pmatrix} u_{yt-1} \\ u_{xt-1} \end{pmatrix}. \tag{17.3.3}$$

As usual, the MA coefficients represent the responses of the variables to u_{yt} and u_{xt} shocks. For instance, θ_{yy} is the response of y_t to a u_{yt} shock one period after it occurred.

The eigenvalues of the MA coefficient matrix in (17.3.3) are

$$\lambda_{1/2} = \frac{\theta_{yy} + \theta_{xx}}{2} \pm \sqrt{\left(\frac{\theta_{yy} + \theta_{xx}}{2}\right)^2 - (\theta_{yy}\theta_{xx} - \theta_{xy}\theta_{yx})}$$

and the process is fundamental if $|\lambda_1|, |\lambda_2| < 1$. Clearly, this condition is satisfied for a wide range of θ_{yy}, θ_{yx}, θ_{xy}, and θ_{xx} values. This example illustrates the rich set of impulse responses that are possible when the additional information in x_t is taken into account. In particular, $\theta_{yy} > 1$ is possible and, hence, a u_{yt} shock can have an effect greater than 1 in the period following its occurrence. For example, for $\theta_{yy} = -\theta_{xx}$ it is easy to construct cases with $\theta_{yy} > 1$ and $|\lambda_1|$, $|\lambda_2| < 1$. Of course, the MA coefficients depend on the properties of x_t and y_t, but, given the large variety of potential x_t variables, it is difficult to exclude any feasible MA coefficient matrices a priori.

If instead of the bivariate joint process z_t only the first component y_t is considered, then the univariate marginal process is known to admit an MA(1) representation as in (17.3.1), with ϕ being a function of θ_{yy}, θ_{yx}, and the covariance matrix parameters of $(u_{yt}, u_{xt})'$ (see Sbrana and Silvestrini (2009) for the precise functional form). As mentioned earlier, in this case there is just one basic nonfundamental univariate MA(1) representation for the given marginal process y_t. In contrast, a continuum of alternative impulse responses is possible if further variables are added. As this discussion shows, a shock having an impact greater than one on y_t one period after its occurrence can also arise from adding further variables to the system under consideration.

Unfortunately, Granger noncausality of x_t for y_t does not solve the problem of distorted impulse responses due to omitted variables completely because there may also be an instantaneous causal relation, for example, if y_t and x_t are contemporaneously correlated. In that case, obtaining orthogonal shocks requires multiplying u_{zt} by some invertible matrix which may affect the impulse responses of y_t. Denoting by Σ_{u_z} the covariance matrix of u_{zt} and by I an identity matrix of suitable dimensions, let

$$Q = \begin{bmatrix} Q_{yy} & Q_{yx} \\ Q_{xy} & Q_{xx} \end{bmatrix}$$

be some invertible matrix partitioned according to the partitioning of z_t such that $Q^{-1}\Sigma_{u_z}Q^{-1\prime} = I$. Then, if $\Theta(L)$ is lower block-triangular,

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Theta_{yy}(L)Q_{yy} & \Theta_{yy}(L)Q_{yx} \\ \Theta_{xy}(L)Q_{yy} + \Theta_{xx}(L)Q_{xy} & \Theta_{xy}(L)Q_{yx} + \Theta_{xx}(L)Q_{xx} \end{bmatrix} w_t$$

is a possible structural fundamental MA representation with orthogonal shocks $w_t = Q^{-1}u_{zt}$. Hence, all shocks in w_t may have nonzero effects on y_t although x_t is not Granger causal for y_t. Thus, fundamentalness is necessary, but not sufficient for valid impulse response analysis, when important variables are omitted from the analysis. Even if x_t is not Granger causal for y_t, it may have to be considered in the impulse response analysis. This problem obviously cannot be solved by allowing for noninvertible MA representations. These considerations show that nonfundamentalness is only part of the problem caused by omitted variables.

Synopsis. The bottom line of this discussion is that there are two ways for overcoming the problem that the fundamental representation of the marginal process of y_t may not adequately reflect the responses to structural shocks. One possibility is to condition on the original set of variables y_t when deriving nonfundamental MA representations of y_t, while the second possibility is to increase the dimension of the process considered. The former possibility can be quite restrictive compared to the latter, because it does not account for instantaneous relations between included and omitted variables. When the basic argument for nonfundamentalness is that the econometrician has not accounted for all relevant information, it makes sense to add the missing information to the system to be considered. This is not to say that small-scale VAR models should never be used. Rather it highlights a potential drawback of fitting small-dimensional VAR models. Avoiding this drawback may call for greater care in how the VAR model is specified and identified. Sections 17.3.4 and 17.3.5 discuss several potential responses to this problem.

This discussion also has implications for the nature of the structural shocks identified by VAR models of the macroeconomy. In structural VAR analysis typical shocks of interest are demand, supply, monetary, fiscal, or technology shocks. Suppose the technology shock is of interest. Clearly in a small-scale VAR model real variables such as output and consumption are not broken down by sector, so an aggregate technology shock can only be thought of as an average technology shock and the implied impulse responses are the responses to such an average shock. In practice, it is rare for a technology shock to affect all sectors of the economy evenly, however; rather a technological innovation occurs in one sector and from there is transmitted to other sectors, eventually affecting the whole economy. Tracing the transmission from one sector to the system may not only paint a different picture of the model dynamics than tracing an average shock, but it also requires sectoral variables to be contained in the system under consideration. Clearly, the correct transmission mechanism cannot be estimated by simply considering both fundamental and nonfundamental shocks in the original VAR model. It requires a larger-dimensional VAR model.

17.3.4 Avoiding Nonfundamentalness by Using Factor-Augmented or Large Bayesian VARs

Avoiding nonfundamental representations in structural impulse response analysis requires the use of VAR models that include a sufficient number of variables. Conventional VAR models include only a small number of variables because otherwise the estimation and specification uncertainty increases so much that the models become uninformative with respect to the relations between the variables involved. This does not mean that there is not a much larger number of variables that one could argue should be included in the model on economic grounds.

How Many Variables Do We Need? An important question is how many variables are required for a structural VAR model to include all the information required to answer the questions of interest. Forni and Gambetti (2014) suggest a formal statistical procedure that allows one to verify whether a given VAR model contains a suitable number of variables. Their proposal is to compute a possibly large number of principal components from the database of potentially relevant economic variables and to test whether these principal components are Granger causal for the smaller set of variables already contained in the VAR model. If none of the principal components is Granger causal for the variables in a given VAR, they conclude that the information in the VAR is sufficient for structural analysis. Otherwise further variables have to be added to the VAR. These variables can be observables or principal components, as would be the case in a FAVAR model. An important limitation of this approach is that it assumes that the factors fully describe the economic agents' information set.

An alternative test of the fundamentalness of a stochastic process has been proposed by Chen, Choi, and Escanciano (2017). Chen et al. focus on VAR processes with non-Gaussian iid structural errors. They prove, under some regularity conditions, that the innovations of the Wold MA representation of the VAR process are martingale difference sequences (mds), if and only if the structural VAR shocks are fundamental. Chen et al.'s test of the mds property of the Wold innovations has an asymptotic $\mathcal{N}(0, 1)$ distribution under the null hypothesis. If the mds null hypothesis is rejected, the test may be repeated after augmenting the original VAR model by additional variables. It is shown by example that adding external estimates of fiscal news to a VAR of fiscal policy shocks may help overcome nonfundamentalness.

Why Large-Scale Models May Not Be the Answer. A common response to the problem of informational deficiencies has been the use of FAVAR models. The imposition of shrinkage restrictions on larger-dimensional VAR models has been another. A third proposal, associated with Forni, Giannone, Lippi, and Reichlin (2009), is to adapt the class of dynamic factor models to allow the construction of structural impulse responses (see Chapter 16).

These alternative models, however, have their own limitations. For example, the structural impulse responses generated by any of these three classes of models are only approximately valid, and the approximation may be poor. In particular, whether factors summarize the information in a large set of variables more accurately than simple aggregation of sectoral information, for example, is not at all clear. Neither form of information aggregation takes into account the economic relations between the disaggregates.

The same type of problem is encountered in large-scale Bayesian VARs (see Chapter 16). For example, Bańbura, Giannone, and Reichlin (2010) use the classical Litterman or Minnesota prior that was derived with the model's forecast accuracy in mind (see also Chapter 5). Priors in large-scale Bayesian VARs serve to compensate for degrees-of-freedom limitations. Hence, they are bound to distort the parameter estimates and also the estimated impulse responses. Caution is called for in interpreting these estimates because unlike in small-dimensional VAR models we cannot evaluate the effect of the prior on the estimates against the benchmark of an unrestricted model.

An even bigger concern than the econometric challenges of working with large-dimensional VAR models is that the information that the econometrician is lacking more often than not is not included in standard macroeconomic databases. In this case, FAVAR models or large-scale BVAR models will not be able to overcome the informational deficiencies of the VAR model. For example, information about the credibility of promises of fiscal reform cannot be captured by simply adding a broad range of macroeconomic indicators to the VAR model. Moreover, to the extent that there are any data at all related to agents' expectations about future fiscal policies from opinion polls or prediction markets, these unconventional data sets tend to be too short for inclusion in a VAR model.

This point is even more apparent when considering the example of a structural VAR model of the global oil market. Expectations about future oil supply disruptions in the Middle East may depend on one's view of the stability of the Saudi monarchy, the resilience of Kurdish fighters to ISIS, the willingness of the Iranian government to compromise on its nuclear ambitions, and numerous other determinants that are inherently unobservable, but nevertheless potentially important for the forward-looking component of the real price of oil. Such expectations cannot possibly be captured by including additional macroeconomic variables in the VAR model.

A more promising approach therefore may be to determine a small number of judiciously chosen additional variables that capture the most important missing information from an economic point of view. This alternative approach is discussed in the next section based on two concrete examples from the literature. We also consider a third example that incorporates forward-looking behavior directly into the structural VAR model. Finally, the reader is referred to Chapter 15 for alternative approaches to augmenting VAR models with external information obtained from survey data or other proxies for non-model-based expectations.

17.3.5 Other Approaches to Dealing with Anticipation

One option for dealing with the informational deficiencies of VAR models is to appeal to the ability of asset markets to aggregate all available information. If asset markets are efficient, asset prices will incorporate all the information available to agents and adding asset prices to the VAR model should help align the information set of the econometrician and the agent.

Example 1: A VAR Model of Anticipated Fiscal Policy. With respect to fiscal foresight, this approach was exploited by Leeper, Walker, and Yang (2013). In the United States, municipal bonds are exempt from federal taxes. The differential tax treatment of municipal and treasury bonds may be used to identify news about tax changes. All else equal, the municipal bond yield spreads allow one to infer the implicit tax rate at which financial investors are indifferent between tax-exempt bonds and taxable bonds. Because bond traders are forward looking, the implicit tax rate predicts future individual income tax rates. If there is news that individual tax rates are expected to increase, for example, investors will drive up the yield on taxable bonds until equilibrium is restored. By comparing the implicit tax rates at different maturities, we may infer shifts in expectations about future taxes. As stressed by Leeper, Walker, and Yang (2013), this approach has several advantages in practice. First, there is no need to specify a priori the period of foresight or for that period to be constant over time. Second, there is no need to specify the functional form of the information flow in the economy. Third, conditioning on the implicit tax rate resolves the nonuniqueness of the MA representation.

Leeper, Walker, and Yang (2013) report that the implicit tax rate variable Granger causes the variables in the fiscal policy VAR model of Blanchard and

Perotti (2002) that we already discussed in Chapter 8, indicating that the latter model is informationally deficient. How important the informational deficiency is, can be shown by revisiting the impulse response analysis. Leeper, Walker, and Yang (2013) propose augmenting Blanchard and Perotti's structural VAR model by the spread variable s_t such that

$$\begin{aligned} u_t^{tax} &= 2.08u_t^{gdp} + w_t^{tax} \\ u_t^{gov} &= c_{21}w_t^{tax} + w_t^{gov} \\ u_t^{gdp} &= -b_{31,0}u_t^{tax} - b_{32,0}u_t^{gov} + w_t^{gdp} \\ u_t^{s} &= -b_{41,0}u_t^{tax} - b_{42,0}u_t^{gov} - b_{43,0}u_t^{gdp} + w_t^{s} \end{aligned}$$

They add the identifying assumption that news contained in the interest rate spread, w_t^s, has no direct effect on current output (u_t^{gdp}), tax revenue (u_t^{tax}), and government spending (u_t^{gov}). The resulting structural VAR model can be used to construct responses both to unanticipated and anticipated tax revenue shocks. Leeper et al. show that their model produces markedly different impulse response estimates from Blanchard and Perotti's model in that anticipated tax increases raise output substantially for about three years, before output begins to decline. There is evidence that agents' foresight extends as far as five years.

As acknowledged by Leeper, Walker, and Yang (2013), there are obvious limitations to using municipal bonds as a measure of anticipated tax changes. First, this approach relies on other factors that affect municipal bonds such as default risk and liquidity risk being negligible. Second, the marginal investor in this market may not be representative of the tax payer. Third, municipal bonds respond to changes in individual income taxes only. Finally, this approach appears specific to the United States. Nevertheless, this example illustrates that there are creative approaches to overcoming informational deficiencies of standard VAR models.

Example 2: A VAR Model of Anticipated Oil Supply Shortfalls. Our second example is the model of the global market for crude oil since 1973 proposed in Kilian and Murphy (2014) (see Chapter 13). Essentially the same model has also been employed by Kilian and Lee (2014). The starting point of Kilian and Murphy's analysis is the observation that there clearly is an important forward-looking component in the real price of oil that cannot be captured by the variables traditionally included in structural oil market models. To resolve this informational deficiency they propose the inclusion of the change in global crude oil inventories as an additional variable in the VAR model.

Their approach is more subtle than that in Leeper, Walker, and Yang (2013) in that Kilian and Murphy (2014) make no attempt to measure expectations directly. Oil price expectations reflect the expected shortfall of future oil supply relative to future oil demand. Kilian and Murphy (2014) exploit the fact that shifts in expectations not already captured by conventional oil market models affect the price of oil in the physical oil market by shifting the demand for crude oil stocks or inventories. To the extent that we can isolate the component of the change in crude oil inventories driven by such demand shifts from the remaining components associated with more conventional oil supply and oil demand shocks, we can infer shocks to expectations without having to quantify these expectations. Kilian and Murphy (2014) propose a set of identifying assumptions motivated by economic theory that is designed to disentangle these shocks and their effect on all model variables simultaneously. This approach explicitly deals with the problem that many of the determinants of oil price expectations are unobservable.

Thus, including the change in crude oil inventories in the model resolves the informational deficiency of the VAR model without having to model expectations. It should be noted that under perfect arbitrage exactly the same information conveyed by the change in inventories will also be contained in the spread of the oil futures price over the spot price of oil (see Alquist and Kilian 2010). Thus, including the oil futures spread in principle would be an alternative way of resolving the model's informational deficiency. In practice, the use of inventory data is preferred (1) because liquid oil futures markets only exist for a small part of the sample, (2) because traders in oil futures markets may not be representative for traders in the physical market for oil, and (3) because there is no need to take a stand on the extent of arbitrage between the physical and the financial market for crude oil when using oil inventory data. Moreover, possible concerns over the quality of the inventory data can be dealt with by comparing the fit of the model against extraneous evidence and by examining the robustness to alternative definitions of the inventory variable (see Kilian and Lee 2014).

It should be noted that including both the oil futures spread and the change in oil inventories in the structural VAR model is not an option. This would induce a singularity in the model to the extent that both variables capture the same information. Although it is conceivable that the oil futures spread may convey additional information not already captured by oil inventories, it can be shown that the oil futures spread has no additional predictive power for the variables included in the Kilian-Murphy model during the subsample for which the oil futures spread is available, suggesting that nothing is lost by excluding it from the model.

Kilian and Murphy (2014) demonstrate that explicitly modeling forward-looking behavior in oil markets makes a substantial difference for the interpretation of several episodes of rising or falling oil prices including 1979, 1986, 1991. Although their approach is specifically motivated by the economic structure of oil markets, it once again illustrates that there are creative solutions to the problem of nonfundamentalness that do not involve the use of large-scale VAR models.

Example 3: A VAR Model of Anticipated Technology Shocks. Not all structural VAR models dealing with forward-looking behavior rely on augmented information sets. A very different approach to modeling forward-looking behavior was proposed by Barsky and Sims (2011). This study focuses on expectations about aggregate technology specifically. They postulate that the log of aggregate technology, a_t, is characterized by a stochastic process driven by two structural shocks. The first shock is the traditional surprise technology shock, which impacts the level of productivity in the same period in which agents observe it. The second shock reflects information about future technology and is defined to be orthogonal to the first shock.[3] The two structural shocks jointly account for all variation in a_t. They are identified as follows:

$$a_t = [B_{11}(L), B_{12}(L)] \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where $B_{12}(0) = 0$ such that only w_{1t} affects current productivity instantaneously, making w_{2t} the future technology shock. Effectively, Barsky and Sims treat a_t as predetermined with respect to the rest of the economy. This identifying assumption leaves a wide range of possible choices for w_{2t}. In practice, w_{2t} is identified as the shock that best explains future movements in $a_{t+1}, \ldots, a_{t+H}$, not accounted for by its own innovation, where H is some finite horizon.

The estimated VAR model includes a total factor productivity series as well as selected macroeconomic aggregates. The variable a_t is ordered first. The procedure is implemented by constructing candidate solutions of the form PQ, where P denotes the lower-triangular Cholesky decomposition of the reduced-form error covariance matrix Σ_u and Q denotes a conformable orthogonal matrix, as in the case of sign-identified VAR models. The ability of a shock to explain future movements of the data is measured in terms of the forecast-error variance decomposition. Because the contribution of the second shock to the forecast error variance of a_t depends only on the second column of B_0^{-1}, Barsky and Sims choose the second column, γ, to solve the optimization problem:

$$\gamma^* = \arg\max_{\gamma} \sum_{h=0}^{H} \omega_{12}(h),$$

subject to the first element of γ being zero and $\gamma'\gamma = 1$. Here $\omega_{ij}(h)$ denotes the share of the forecast error variance of variable i attributable to structural shock j at horizon h expressed in terms of the structural parameters of the

[3] Barsky and Sims refer to this shock as a news shock, following a terminology common in the recent macroeconomic literature. This is somewhat misleading in that news shocks have traditionally been defined as unexpected changes to observed aggregates (see Chapter 7). Rather the second shock captures expected changes in future productivity.

model (see Chapter 4). The criterion function can be maximized, for example, by searching the space of possible subrotation matrices much like in the literature on sign-identified VAR models (see Chapter 13). One potential limitation of this solution method is that it generates the best possible case for expectations shocks rather than the most likely case. To the extent that the fraction of the forecast error variance of real GDP is small even in the best possible case, of course, this procedure will generate informative results.

Example 4: VAR Models with External Policy Shocks. A fourth example of how to deal with forward-looking expectations involves the use of external measures of policy shocks, as discussed in Chapter 15. For example, the Romer and Romer (2004) measure of monetary policy shocks is derived on the basis of internal forecasts of the Federal Reserve Board that reflect more information than can be captured by a VAR model. Likewise measures of exogenous policy shifts based on changes in interest rate futures prices around FOMC dates allow for richer information sets to be added to the VAR model.

17.4 Conclusions

One argument against using conventional structural VAR models for macroeconometric analysis is that they capture only fundamental shocks. These are shocks that can be recovered from past and present observed variables. In turn, shocks that are not recoverable from present and past observations are called nonfundamental. The fact that economic agents use additional information in decision-making that is not available to, or at least not used by the econometrician in specifying the VAR model, is a typical argument in favor of nonfundamental shocks. Two main approaches have been proposed to address the nonfundamentalness problem. First, in fundamental MA representations all roots of the determinantal polynomial obtained from the MA operator are outside the complex unit circle. Thus, allowing for roots inside the unit circle produces nonfundamental representations and shocks. The second response to the nonfundamentalness problem is to augment the information set by including more variables in the VAR model. Both approaches are plausible solutions to the nonfundamentalness problem, but they also have their limitations.

Allowing for MA roots inside the unit circle is plausible if the underlying economic model has this feature. This approach will not necessarily capture the true impulse responses, however, if the nonfundamentalness arises from omitted variables in the VAR model. Moreover, allowing for MA roots inside the unit circle complicates the identification of the structural shocks, unless economic theory provides a full set of identifying restrictions, which is rarely the case in practice.

Viewing nonfundamentalness as an omitted variables problem is attractive, given that nonfundamentalness typically is caused by the econometrician not using all relevant information in specifying the VAR model. Adding the missing information to the model is a natural response to the problem. There are two complications, however.

First, adding all information that may be important in agents' decision-making involves a large number of variables, resulting in the well-known curse of dimensionality. The latter problem can be tackled in different ways. One proposal is to use structural FAVAR models. We observed that, while these models are useful for integrating large panels of variables within the same VAR model, it is not straightforward to extract the structural impulse responses of interest from these models. Similar concerns apply to the use of large-scale Bayesian structural VAR models.

Second, the information relevant to the decisions of economic agents may not be captured by the variables contained in standard databases. Clearly, neither FAVAR models nor large-scale Bayesian VAR models are designed to address this second concern.

In short, each of the solutions to the problem of nonfundamentalness proposed in the literature is beset with its own limitations. In the end, any proposal for overcoming the curse of dimensionality, which results from including a large number of variables, can be criticized on the grounds that it may induce distortions in impulse responses. A more promising approach therefore may be to identify a small number of variables on economic grounds that reflect the relevant information for a particular problem at hand and including those in the model. Although it may not be easy to agree on the most important variables in general, in specific cases economic theory may provide some guidance. We discussed four examples in the empirical literature that propose creative solutions to the problem of nonfundamentalness.

One final point to keep in mind is made by Sims (2012). He shows by example that nonfundamentalness leads to distortions in standard VAR analysis, but that these distortions may be small in practice. For example, impulse response functions obtained from standard structural VAR models may be very similar to the true responses even if the true shocks are nonfundamental. A similar point is also made in Beaudry and Portier (2014). These authors provide a detailed discussion of the nonfundamentalness problem in the context of the news shocks defined in their work.