FISCAL FORESIGHT AND INFORMATION FLOWS

ERIC M. LEEPER, TODD B. WALKER, AND SHU-CHUN SUSAN YANG

Abstract. News—or foresight—about future economic fundamentals can create rational expectations equilibria with non-fundamental representations that pose substantial challenges to econometric efforts to recover the structural shocks to which economic agents react. Using tax policies as a leading example of foresight, simple theory makes transparent the economic behavior and information structures that generate non-fundamental equilibria. Econometric analyses that fail to model foresight will obtain biased estimates of output multipliers for taxes; biases are quantitatively important when two canonical theoretical models are taken as data generating processes. Both the nature of equilibria and the inferences about the effects of anticipated tax changes hinge critically on hypothesized information flows. Different methods for extracting or hypothesizing the information flows are discussed with an emphasis on how differential U.S. federal tax treatment of municipal and treasury bonds embeds news about future taxes in bond yield spreads. Including that measure of tax news in identified VARs produces substantially different inferences about the macroeconomic impacts of anticipated taxes.

Keywords: news, anticipated taxes, non-fundamental representation, identified VARs
JEL Codes: C5, E62, H30

1. Introduction

A venerable tradition, often traced to Pigou (1927), ascribes a significant role in aggregate fluctuations to economic decision makers’ responses to expectations about not-yet-realized economic fundamentals. That tradition finds voice in a recent surge of interest in the economic consequences of news—or foresight. Recent work explores how news affects the predictions of standard theories, seeks evidence of the impacts of news in time series data, and estimates dynamic stochastic general equilibrium models to quantify the relative importance of anticipated and unanticipated “shocks” to fundamentals.

Existing work typically posits a particular stochastic process for news, grounded in neither theory nor empirics. That process determines the economy’s information flows and, in

Date: April 10, 2012. Department of Economics, Indiana University, Monash University and NBER, eleeper@indiana.edu; Department of Economics, Indiana University, walkertb@indiana.edu; International Monetary Fund, SYang@imf.org. Walker acknowledges support from NSF grant SES-0962221. Yang thanks Academia Sinica for support in the early stages of this research. We also acknowledge comments by Troy Davig, Mike Dotsey, Jesús Fernández-Villaverde, Dale Henderson, Beth Klee, Karel Mertens, Jim Nason, Ricardo Nunes, Valerie Ramey, Morten Ravn, Chris Sims, and participants at many conferences and presentations. Joonyoung Hur provided excellent research assistance. We are particularly grateful to Harald Uhlig and four anonymous referees for helpful comments. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.
a rational expectations equilibrium, agents’ expectations. Given the prominent role of expectations in the news literature, it is remarkable that existing work does not systematically examine how the specification of information flows affects the nature of equilibrium and the connection of theory to data. This paper addresses that gap.

For several reasons we focus on how to identify and quantify the impacts of foreseen “shocks” to taxes. First, few economic phenomena provide economic agents with such clear signals about how important margins will change in the future: foresight is endemic to tax policy. Second, an institutional structure governs information flows about taxes: the process of changing taxes entails two kinds of lags—the inside lag, between when new tax law is initially proposed and when it is passed, and the outside lag, between when the legislation is signed into law and when it is implemented. That institutional structure informs the nature of tax information flows. Third, differential U.S. tax treatment of municipal and treasury bonds leads to a direct measure of tax news that offers a potential solution to modeling tax foresight. Such measures are scarce for news about nonpolicy fundamentals like total factor productivity. Despite the paper’s focus on taxes, one of its key messages—that hypothesized information flows are critical to determining the impacts of news—extends immediately to other contexts.¹

Fiscal foresight poses a challenge to econometric analyses of fiscal policy because it generates an equilibrium with a non-fundamental moving average representation. Information sets of economic agents and the econometrician tend to be misaligned, with agents basing their choices on more information than the econometrician possesses. Structural shocks to tax policy, then, cannot be recovered from current and past fiscal data, a central assumption of conventional econometric methods. Instead, conventional methods can lead the econometrician to label as “tax shocks” objects that are linear combinations of all the exogenous disturbances at various leads and lags.²

This paper builds on and extends Hansen and Sargent’s (1991b) general characterization of the implications of environments in which the history of innovations in a vector autoregression does not equal the history of information that agents observe. First, we go beyond treating invertibility as a 0–1 proposition by assessing the quantitative importance of failing to model foresight in two workhorse macroeconomic models. Second, we offer a compelling economic example—tax foresight—that makes clear that non-fundamentalness and its consequences affect answers to substantive macroeconomic questions. Most importantly, we ground non-fundamentalness in economic theory, which points toward an empirical line of attack that we pursue. Both Hansen and Sargent (1991b) and Fernández-Villaverde, Rubio-Ramírez, ¹In addition to taxes, studies have examined news about a wide range of fundamentals, including total factor and investment-specific productivity [Beaudry and Portier (2006), Christiano, Ilut, Motto, and Rostagno (2008), Jaimovich and Rebelo (2009), Schmitt-Grohé and Uribe (2008), Fujiwara, Hirose, and Shintani (2011)]; government military spending run ups [Fisher and Peters (2009), Ramey (2011)]; phased-in government infrastructure spending [Leeper, Walker, and Yang (2010)]; announcements of interest-rate paths by inflation-targeting central banks [Blattner, Catenaro, Ehrmann, Strauch, and Turunen (2008), Laséen and Svensson (2011)]. All of these applications lend themselves to the analysis that we conduct.

Sargent, and Watson (2007) have been read primarily as cautionary notes, in large part because they point to a serious problem but not to a way forward.

No consensus exists on how to handle tax foresight, a fact that is underscored by the diverse empirical findings in the literature. Research concludes that an anticipated cut in taxes may have little or no effect [Poterba (1988), Blanchard and Perotti (2002), Romer and Romer (2010)], may be mildly expansionary in the short run [Mountford and Uhlig (2009)], or may be strongly contractionary in the short run [House and Shapiro (2006), Mertens and Ravn (2011)]. By using different measures of tax news, these studies implicitly posit different tax information flows, which, as we show, can produce strikingly different inferences about the effects of anticipated tax changes.

The paper has three major parts:

(1) A simple analytical example makes precise how foresight and optimizing behavior create equilibria with non-fundamental moving average representations. The example makes the source of non-fundamentalness transparent: it arises as a natural by-product of the fact that agents’ optimal intertemporal decisions discount future tax obligations. Although private agents discount tax rates in the usual way, they discount recent tax news more heavily than past news because with foresight the recent news informs about taxes in the more distant future. The econometrician, in contrast, discounts in the usual way, down weighting older news relative to recent news. Agents and the econometrician employ different discounting patterns because the econometrician’s information set lags the agents’.

(2) Simple analytics reveal the source of non-fundamentalness, but do not shed light on whether it matters in practice. Using two canonical dynamic stochastic general equilibrium models—Chari, Kehoe and McGrattan’s (2008) real business cycle model and Smets and Wouters’ (2003; 2007) new Keynesian model—as data generating processes, we quantify the inference errors an econometrician might make by failing to model foresight. We tie those errors to alternative, empirically motivated specifications of tax news processes—information flows that distinguish between the “inside” and “outside” lags associated with tax policies. Estimates of tax multipliers can be off by hundreds of percent and even be of the wrong sign. Biases can be positive or negative, but the econometrician tends to underestimate the effects of foresight over longer horizons.

(3) We discuss several lines of attack that offer a way forward in dealing with non-fundamental equilibria. We show that seemingly unrelated approaches [e.g., the narrative approach of Ramey (2007) and Romer and Romer (2010) and the dynamic stochastic general equilibrium approach of Schmitt-Grohé and Uribe (2008)], are in fact solving the problems associated with foresight in a similar fashion—by expanding the information set of the econometrician. We offer a fresh set of results that exploits the differential treatment in the U.S. tax code that exempts municipal bonds from federal income tax to extract news about future tax changes from the spread between municipal and treasury bond yields.
2. Analytical Example

This section introduces fiscal foresight into a simple economic environment where the econometric issues can be expositored analytically. Results and conclusions reached in the simple exposition extend to more general setups, as section 2.2 discusses.

Consider a standard growth model with a representative household that maximizes expected log utility, \( E_0 \sum_{t=0}^{\infty} \beta^t \log(C_t) \), subject to \( C_t + K_t + T_t \leq A_t K_{t-1}^\alpha \), where \( C_t, K_t, Y_t, \) and \( T_t \) denote time-\( t \) consumption, capital, output, and lump-sum taxes, respectively, and \( A_t \) is an exogenous technology shock. As usual, \( 0 < \alpha < 1 \) and \( 0 < \beta < 1 \). The government sets the tax rate according to a time-invariant rule and adjusts lump-sum transfers to satisfy the constraint, \( T_t = \tau_t Y_t \). Government spending is identically zero. Labor is supplied inelastically which, as section 3.3 shows, understates the problems that foresight creates.

The equilibrium conditions are well known and given by
\[
\frac{1}{C_t} = \alpha \beta E_t \left[ (1 - \tau_{t+1}) \frac{1}{C_{t+1}} \frac{Y_{t+1}}{K_t} \right]
\] (1)
\[
C_t + K_t = Y_t = A_t K_{t-1}^\alpha.
\] (2)

Let \( A \) and \( \tau \) denote the steady state values of technology and the tax rate. The steady state capital stock is \( K = [\alpha \beta (1 - \tau) A]^{1/(1 - \alpha)} \). Let lower case letters denote percentage deviations from steady state values, \( k_t = \log(K_t) - \log(K) \), \( a_t = \log(A_t) - \log(A) \), and \( \hat{\tau}_t = \log(\tau_t) - \log(\tau) \). Log linearizing (1)–(2) yields an equilibrium that is characterized by a second-order difference equation in capital
\[
E_t k_{t+1} - (\theta^{-1} + \alpha)k_t + \alpha \theta^{-1} k_{t-1} = E_t [a_{t+1} - \theta^{-1} a_t] + \left\{ \theta^{-1} (1 - \theta) \left( \frac{\tau}{1 - \tau} \right) \right\} E_t \hat{\tau}_{t+1},
\] (3)

where \( \theta = \alpha \beta (1 - \tau) \) is a particularly important constant in the analysis. Assuming an \( i.i.d. \) technology shock, the solution to (3) is
\[
k_t = \alpha k_{t-1} + a_t - (1 - \theta) \left( \frac{\tau}{1 - \tau} \right) \sum_{i=0}^{\infty} \theta^i E_t \hat{\tau}_{t+i+1}.
\] (4)


To model foresight, we must specify how news about taxes signals future tax rates. For many of the points we wish to make, it suffices to assume that tax information flows take a particularly simple form: agents at \( t \) receive a signal that tells them exactly what tax rate they will face in period \( t + q \). In later sections we will relax this assumption and posit more sophisticated rules for tax rates. The tax rule is \( \tau_t = \hat{\tau} e^{\varepsilon_{\tau,t-q}} \), or in log-linearized form
\[
\hat{\tau}_t = \varepsilon_{\tau,t-q}.
\] (5)

Assume the technology and tax shocks are \( i.i.d. \) and the representative agent’s information set at date \( t \) consists of variables dated \( t \) and earlier, including the shocks, \( \{ \varepsilon_{A,t}, \varepsilon_{\tau,t} \} \). Given the tax rule in (5), this implies that at \( t \) the agent has (perfect) knowledge of \( \{ \hat{\tau}_{t+q}, \hat{\tau}_{t+q-1}, \ldots \} \).
Using the information flows in the tax rule to solve for expected tax rates in (4) for various degrees of fiscal foresight yields the following equilibrium dynamics.

$q = 0$ implies:

\[ k_t = \alpha k_{t-1} + \varepsilon_{A,t} \]  

(6)

$q = 1$ implies:

\[ k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa \varepsilon_{\tau,t} \]  

(7)

$q = 2$ implies:

\[ k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa \left\{ \varepsilon_{\tau,t-1} + \theta \varepsilon_{\tau,t} \right\} \]  

(8)

$q = 3$ implies:

\[ k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa \left\{ \varepsilon_{\tau,t-2} + \theta \varepsilon_{\tau,t-1} + \theta^2 \varepsilon_{\tau,t} \right\} \]  

(9)

where \( \kappa = (1 - \theta)(\tau/(1 - \tau)) \).

If there is no foresight, \( q = 0 \), we get the usual result that \textit{i.i.d.} shocks to tax rates have no effect on capital accumulation. When there is some degree of tax foresight \((q > 0)\), rational agents will adjust capital contemporaneously to yield the unusual result that even serially uncorrelated tax hikes reduce capital accumulation. Fiscal foresight manifests in the additional moving average terms present in the equilibrium representation. The number of moving average terms increases with the foresight horizon.

A striking, though seemingly perverse, implication of (8) and (9) is that more recent news is discounted (by \( \theta = \alpha \beta (1 - \tau) < 1 \)) relative to older news. This is because with two-quarter foresight, \( \varepsilon_{\tau,t-1} \) affects \( \hat{\tau}_{t+1} \), while \( \varepsilon_{\tau,t} \) affects \( \hat{\tau}_{t+2} \), so the news that affects tax rates farther into the future receives the heaviest discount. While tax \textit{rates} are discounted in the usual way, tax \textit{news} is discounted in reverse order. This difference in discounting between tax rates and tax news stems from optimizing behavior and underlies the econometric problems that foresight creates.

2.1. The Econometrics of Foresight. The moving average terms that foresight produces pose challenges for conducting econometric inference. Conventional econometric analyses, such as those using identified vector autoregressions (VARs), can draw erroneous conclusions. Errors arise because models with foresight may imply that the information set of private agents is larger than the econometrician’s.

An econometrician who estimating an identified VAR aims to condition on the same information set as the economic agents to recover the structural shocks \( \{ \varepsilon_{\tau,t-j} \}_{j=0}^{\infty} \). Typically, this is achieved by conditioning the VAR estimates on current and past observable variables. Consider the univariate case of conditioning on current and past capital, \( \{ k_{t-j} \}_{j=0}^{\infty} \), and suppose that agents have two quarters of foresight. Using lag operators (\( i.e., \ L^s x_t = x_{t-s} \)), (8) may be written as

\[ (1 - \alpha L) k_t = -\kappa (L + \theta) \varepsilon_{\tau,t} \]  

(10)
Will the econometrician’s conditioning set, current and past capital, span the same space as the agents’, current and past structural shocks?3 The answer depends on whether \( \{\varepsilon_{\tau,t-j}\}_{j=0}^{\infty} \) is fundamental for \( \{k_{t-j}\}_{j=0}^{\infty} \), using the terminology of Rozanov (1967). Fundamentalness requires the equilibrium process to be invertible in current and past \( k_t \), so that

\[
\left[ 1 - \alpha L \over 1 + \theta^{-1} L \right] k_t
\]

is a convergent sequence. If \( |\theta| > 1 \) this condition holds and \( \{k_{t-j}\}_{j=0}^{\infty} \) spans the same space as \( \{\varepsilon_{\tau,t-j}\}_{j=0}^{\infty} \). But a unique saddlepath solution requires \( |\theta| < 1 \). Therefore, \( \{\varepsilon_{\tau,t-j}\}_{j=0}^{\infty} \) is not fundamental for \( \{k_{t-j}\}_{j=0}^{\infty} \).

To determine the econometrician’s information set, we derive the Wold representation for \( k_t \) from the one-step-ahead forecast errors associated with predicting \( k_t \) conditional only on its past values. This representation emerges from flipping the root of the moving average representation from inside the unit circle to outside the unit circle using the Blaschke factor, \([ (L + \theta)/(1 + \theta L) ]\) [see Hansen and Sargent (1991b) or Lippi and Reichlin (1994)]. The Wold representation for capital is

\[
(1 - \alpha L)k_t = -\kappa (L + \theta) \left[ 1 + \frac{\theta L}{L + \theta} \right] \varepsilon_{\tau,t}
\]

\[
= -\kappa \left( L + \theta \right) \varepsilon^*_{\tau,t}
\]

\[
= -\kappa \left\{ \theta \varepsilon^*_{\tau,t-1} + \varepsilon^*_{\tau,t} \right\}.
\]

By observing current and past capital, the econometrician recovers current and past \( \varepsilon^*_\tau \), rather than the news that private agents observe, current and past \( \varepsilon_\tau \). The econometrician’s innovations are the statistical shocks associated with estimating the autoregressive representation; those shocks turn out to represent information that is mostly “old news” to the agents of the economy. Fundamental shocks map into the econometrician’s shocks as

\[
\varepsilon^*_{\tau,t} = \left[ \frac{L + \theta}{1 + \theta L} \right] \varepsilon_{\tau,t} = (L + \theta) \sum_{j=0}^{\infty} -\theta^j \varepsilon_{\tau,t-j}
\]

\[
= \theta \varepsilon_{\tau,t} + (1 - \theta^2) \varepsilon_{\tau,t-1} - \theta (1 - \theta^2) \varepsilon_{\tau,t-2} + \theta^2 (1 - \theta^2) \varepsilon_{\tau,t-3} + \cdots
\]

This mapping shows that what the econometrician recovers as the tax innovation at time \( t \), \( \varepsilon^*_{\tau,t} \), is actually a discounted sum of the tax news observed by the agents at date \( t \) and earlier.

An econometrician who ignores foresight will discount the innovations incorrectly. In the econometrician’s representation, yesterday’s innovation has less effect than today’s innovation, as the terms \( \theta \varepsilon^*_{\tau,t-1} + \varepsilon^*_\tau \) in (11) show. Agents with foresight, in contrast, discount news according to \( \varepsilon_{\tau,t-1} + \theta \varepsilon_{\tau,t} \), as in (8), because yesterday’s news has a larger effect on capital accumulation than today’s news. Differences in discounting patterns applied by the econometrician and the agents lead to a variety of econometric problems.

---

3More specifically, the information sets are equivalent if the the Hilbert space generated by \( \{k_{t-j}\}_{j=0}^{\infty} \) is equivalent (in mean-square norm) to the Hilbert space generated by \( \{\varepsilon_{\tau,t-j}\}_{j=0}^{\infty} \).
By employing VAR analysis and not modeling foresight, the econometrician has conditioned on a smaller information set. The extent to which private agents condition on information that is not captured by current and past variables in the econometrician’s information set determines the error associated with the VAR. This error can be mapped directly into the $\theta$ parameter that governs the non-invertibility of the equilibrium moving-average representation. The variance of the one-step-ahead forecast error for the agent is

$$
E[(k_{t+1} - E[k_{t+1}|\varepsilon^t])^2] = E \left[ \left( \frac{-\kappa(L + \theta)}{1 - \alpha L} \varepsilon_{\tau,t+1} - L^{-1} \left[ -\frac{\kappa(L + \theta)}{1 - \alpha L} + \kappa \theta \right] \varepsilon_{\tau,t} \right)^2 \right] = (\kappa \theta)^2 \sigma_{\tau}^2
$$

where $\varepsilon^t$ denotes current and past $\varepsilon$. For the econometrician’s information set, the variance of the forecast error is

$$
E[(k_{t+1} - E[k_{t+1}|k^t])^2] = E \left[ \left( \frac{-\kappa(L + \theta)}{1 - \alpha L} \varepsilon_{\tau,t+1} - L^{-1} \left[ -\frac{\kappa(1 + \theta L)}{1 - \alpha L} + \kappa \right] \left[ \frac{L + \theta}{1 + \theta L} \right] \varepsilon_{\tau,t} \right)^2 \right] = \kappa^2 \sigma_{\tau}^2
$$

The ratio of (13) to (14) is $\theta^2$. As $\theta^2$ approaches unity (zero), the difference between the agent and econometrician’s information sets gets smaller (larger). If $\theta$ is greater than or equal to 1, the representation for capital becomes fundamental with respect to $\varepsilon_{\tau,t}$ and the variances of the forecast errors (13) and (14) coincide.

To examine the importance of the information discrepancies in this model, we plot impulse response functions conditioning on the agents’ and econometrician’s information sets. Impulse response functions are widely used to convey how agents respond to innovations, but response functions based on the econometrician’s information set will not capture these responses. Consider the impulse response functions generated by (8) and (11). Figure 1a plots the responses of capital assuming two quarters of foresight (with $\alpha = 0.36, \beta = 0.99, \tau = 0.25, \sigma_{\tau}^2 = 1$). With foresight, agents know exactly when the innovation in fiscal policy translates into changes in the tax rate. This creates the sharp decline in capital one quarter after the news arrives and before the tax rate changes, as the dotted-dashed line indicates. The econometrician’s VAR, though, discounts the innovations incorrectly and reports that the biggest decline in capital occurs on impact, suggesting that foresight does not exist (solid line). The difference between the response functions can be quite dramatic, especially at short horizons.

Figure 1a shows that the econometrician will infer that the tax shock is unanticipated. Of course, not all shocks that affect fiscal policy are known several quarters in advance. Consider a tax rate process, $\hat{\tau}_t = e_{\tau,t}^u + \varepsilon_{\tau,t-q}$, that allows for both anticipated ($\varepsilon_{\tau}$) and unanticipated ($e_{\tau}^u$) shocks at time $t$. If these shocks are orthogonal at all leads and lags, then the equilibrium dynamics of (3) will not change because i.i.d. tax shocks will not alter the dynamics of capital. An econometrician who does not account for foresight will attribute all of the dynamics associated with the anticipated component of the tax rate to the unanticipated component. This suggests that researchers interested in the dynamic effects of fiscal policy—whether the interest is in anticipated or unanticipated changes in policy—must explicitly account for foresight to avoid spurious conclusions.
FISCAL FORESIGHT AND INFORMATION FLOWS

Figure 1: Responses of Capital to Tax Increase with $\alpha = 0.36, \beta = 0.99, \tau = 0.25$. Figure 1a plots the response of (13) and (14). Figure 1b plots the response to the VAR $\left(\tau_t, k_t\right)'$. Both figures assume two quarters of foresight.

Conditioning on more variables will not always lead to better inference. In the case of two-quarter foresight, suppose the econometrician estimates a VAR that includes the tax rate and the capital stock as observables

$$\hat{\tau}_t = \left[ \begin{array}{c} \frac{L^2}{-\kappa(L+\theta)} & 0 \\ \frac{1}{1-\alpha L} & 1 \end{array} \right] \varepsilon_{\tau,t}$$

$$x_t = H(L)\varepsilon_t.$$  \hspace{1cm} (15)

A necessary condition for $\varepsilon_t$ to be a fundamental for $x_t$ is that the determinant of $H(z)$ be analytic with no zeros inside the unit circle. Foresight creates a zero inside the unit circle (at $z = 0$), implying that the information set generated by $\{x_t, x_{t-1}, x_{t-2}, \ldots\}$ is smaller than the information set generated by $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$.

The Wold representation for (15) is obtained by finding Blaschke matrices $B(L)$ and orthonormal matrices $W, \tilde{W}$ that do not alter the covariance generating function of $x_t$, but “flip” the zeros outside of the unit circle. To do this we search for a $B(z)$ and $W, \tilde{W}$ that satisfy $B(z)B(z^{-1})' = I$ and $WW' = I, \tilde{W}\tilde{W}' = I$, and produces innovations that span the space generated by $\{x_t, x_{t-1}, x_{t-2}, \ldots\}$. Following appendix A of Townsend (1983), the first step in the algorithm is to evaluate $H(z)$ at $z = 0$, and postmultiply by $W$ so as to put the zeros in the first column of the product matrix. The remaining columns of $W$ can be constructed using a Gram-Schmidt orthogonalization procedure. The orthonormal $W$ matrix ensures that the representation remains causal, preserving the assumption that the econometrician does not observe future values of the variables. Postmultiplying by $B(L)$ flips the zero outside of the unit circle. With two zeros inside the unit circle for (15), one must repeat this exercise (find an orthonormal matrix $\tilde{W}$ that aligns the zeros in the first column, etc.). Proceeding in this fashion delivers the representation.
\[
\begin{bmatrix}
\hat{\tau}_t \\
k_t
\end{bmatrix} = \begin{bmatrix}
L^2 & 0 \\
-k(L+\theta) & 1-\alpha L
\end{bmatrix} W B(L) \tilde{W} B(L) B(L^{-1}) \tilde{W}' B(L^{-1}) W' \begin{bmatrix}
\varepsilon_{\tau,t} \\
\varepsilon_{A,t}
\end{bmatrix}
\]

\[
x_t = \mathcal{H}^*(L) \tilde{x}^* + \mathcal{E}^*_t
\]

where

\[
W = \begin{bmatrix}
\frac{1}{\sqrt{1+\theta \kappa^2}} & \frac{-\kappa \theta}{\sqrt{1+\theta \kappa^2}} \\
\frac{-\kappa \theta}{\sqrt{1+\theta \kappa^2}} & \frac{1}{\sqrt{1+\theta \kappa^2}}
\end{bmatrix} ; \tilde{W} = \begin{bmatrix}
\Delta(1 + \kappa^2 \theta^2) & -\Delta \kappa \\
\Delta(1 + \kappa^2 \theta^2) & \Delta \kappa
\end{bmatrix} ; B(L) = \begin{bmatrix}
L^{-1} & 0 \\
0 & 1
\end{bmatrix}
\]

and \( \Delta = [(1 + \kappa^2 \theta^2)^2 + \kappa^2]^{-1/2} \).

Now the econometric problems are more severe. First, the econometrician who proceeds with VAR analysis using (16) will likely obtain an impulse response function in which foresight does not appear to exist in the data. Figure 1b depicts the response of capital to a tax increase for the agent (dotted-dashed line) and econometrician as the variance of the technology shock decreases from 1 to 0.01. Conditioning on the econometrician’s information set, the path of capital is flat when \( \sigma_a^2 = \sigma^2 = 1 \). In theory, unanticipated i.i.d. capital tax shocks have no effect on the economy, so based on the flat response of capital, an econometrician will infer that the effects of fiscal policy are limited to unanticipated components only. By not modeling foresight, this example achieves a “self-fulfilling prophesy” and wrongly concludes that foresight is not an issue.\(^4\)

Second, as the variance of the tax shock increases relative to the technology shock, the errors associated with foresight become more pronounced. Figure 1b shows that the initial response of capital to a one-standard-deviation increase in the tax shock increases from 0 to 0.12 as \( \sigma_a^2 \) decreases from 1 to 0.01, so that an anticipated tax increase could be estimated to have no effect or a positive effect on capital and output.

Existing empirical work reports a diverse set of inferences about the effects of an anticipated tax increase on output. Figures 1a and 1b demonstrate that even this simple model can deliver diverse results that depend on the underlying information flows.

Finally, all conditional statistics reported by the econometrician will be misspecified. Consider the variance decompositions that Hansen and Sargent (1991b) emphasize. Let

\[
E(x_t - E_{t-j}^* x_t)(x_t - E_{t-j}^* x_t)' = \sum_{k=0}^{j-1} \mathcal{H}^*_k \Sigma^* \mathcal{H}^*_k
\]

denote the \( j \)-step ahead prediction error variance associated with the econometrician’s information set, where \( \Sigma^* \) is the variance-covariance matrix associated with \( (\varepsilon_{\tau,t}^*, \varepsilon_{A,t}^*)' \). Like impulse response functions, variance decompositions are derived using conditional expectations, so the discrepancy in the information sets implies the coefficients generated by \( \mathcal{H}^*(L) \)

\(^4\)With this simple form of foresight, an econometrician who estimates a VAR in \( (\hat{\tau}_{t+q}, k_t) \) will recover the true shocks. But more sophisticated information flows, as in later sections, or empirically plausible tax rules, as in Leeper, Plante, and Traum (2010), preclude that easy fix.
will misallocate the variance across the structural shocks.\textsuperscript{5} Figure 1b suggests that the econometrician will treat the tax shock as nearly \textit{i.i.d.} and infer that none of the variation in capital (and hence output) can be attributed to tax innovations; all of the variation will be attributed to the technology shock. This inference holds even if, in fact, the tax shock explained nearly all of the variation in capital (for example, when the variance of the technology shock, $\sigma_2^2$, is arbitrarily small).

We derive further implications of foresight in the online appendix, where we show that Granger causality tests and tests of economic theory, such as tests of present value restrictions, will be misspecified in the presence of foresight. Errors associated with ignoring foresight can be quite large.

\textbf{2.2. Generalizations.} The previous example assumes an \textit{i.i.d.} tax shock, but the difficulties associated with foresight extend to more general setups. Suppose the stationary tax rate follows $\hat{\tau}_t = C(L)L^q\varepsilon_{\tau,t}$, where $C(L)$ is a polynomial in the lag operator $L$ and $q$ is the degree of foresight. The only restriction placed on $C(L)$ is that the corresponding coefficients are square summable, which allows for \textit{any} serial correlation pattern. Agents guess that the law of motion for capital is given by a square summable linear combination of tax and technology shocks, $k_t = F(L)\varepsilon_{\tau,t} + G(L)\varepsilon_{\omega,t}$, as Whiteman (1983) shows. Focusing on tax shocks only and substituting this guess into the difference equation for capital in (3) yields

$$\theta L^{-1}[F(L) - F_0]\varepsilon_{\tau,t} = (1 + \alpha \theta)F(L)\varepsilon_{\tau,t} + \alpha LF(L)\varepsilon_{\tau,t} = \left\{ (1 - \theta) \left( \frac{\tau}{1 - \tau} \right) \right\} \hat{E}_{t+1} \hat{\tau}_{t+1}$$

where the Weiner-Kolmogorov formula is used to take expectations (i.e., $E_t x_{t+1} = L^{-1}[D(L) - D_0]x_{t+4}$), and $\theta = \alpha \beta (1 - \tau)$. Uniqueness of the rational expectations equilibrium requires $|\theta| < 1$, where the equilibrium $F(L)\varepsilon_{\tau,t}$ for $q$ degrees of foresight is given by

$$F(L)\varepsilon_{\tau,t} = - \left[ \frac{\kappa [L^q C(L) - \theta^q C(\theta)]}{(1 - \alpha L)(L - \theta)} \right] \varepsilon_{\tau,t}.$$  

(17)

This equation makes plain how foresight impinges on optimal capital accumulation for any choice of $C(L)$. Whenever $q \geq 2$, the equilibrium contains moving average components even when $C(L)$ is purely autoregressive. This representation suggests that it is straightforward to construct impulse response functions that take a wide range of shapes (including hump-shaped), for which the dynamic equation for capital continues to be non-invertible in current and past $k_t$. For example, setting $C(L) = (1 - \rho_1 L - \rho_2 L^2)^{-1}$ and assuming two quarters of foresight ($q = 2$) implies that the tax shocks $\varepsilon_{\tau,t}$ are non-fundamental for $k_t$ if $\theta < (1 + \rho_1)^{-1}$. Because the condition for a non-fundamental moving average representation is independent of $\rho_2$, impulse response functions of non-fundamental moving average representations can adopt many forms.

The logic that leads foresight to produce equilibria with non-fundamental moving-average representations extends to a large class of models. Consider the generic multivariate rational expectations model

$$\Gamma_0 y_t = \Gamma_1 y_{t-1} + \Psi z_t + \Pi \eta_t,$$

(18)

\textsuperscript{5}This result holds even though the statistical shocks of the VAR remain uncorrelated. Orthogonality of the Blaschke and $W$ matrices ($B(L)B(L^{-1}) = I$ and $WW' = \tilde{W}W' = I$) implies that the unconditional second moments of the VAR system remain the same, but the conditional moments will be different.
where \( y_t \) is an \( n \times 1 \) vector of endogenous variables, \( z_t \) is an \( m \times 1 \) vector of exogenous random shocks, \( \eta \) is a \( k \times 1 \) vector of expectation errors, which satisfy \( E_t \eta_{t+1} = 0 \) for all \( t \). \( \Gamma_0 \) and \( \Gamma_1 \) are \( n \times n \) coefficient matrices, along with \( \Psi (n \times m) \) and \( \Pi (n \times k) \). Klein (2000) and Sims (2002) use a generalized Schur decomposition of \( \Gamma_0 \) and \( \Gamma_1 \) to show that there exist matrices such that

\[
\begin{bmatrix}
\Lambda_{11} & \Lambda_{12} \\
0 & \Lambda_{22}
\end{bmatrix}
\begin{bmatrix}
w_{1,t} \\
w_{2,t}
\end{bmatrix}
= 
\begin{bmatrix}
\Omega_{11} & \Omega_{12} \\
0 & \Omega_{22}
\end{bmatrix}
\begin{bmatrix}
w_{1,t-1} \\
w_{2,t-1}
\end{bmatrix}
+ 
\begin{bmatrix}
Q_1 \\
Q_2
\end{bmatrix}
(\Psi z_t + \Pi \eta_t) \tag{19}
\]

The system is partitioned so that the generalized eigenvalues imply an explosive path for \( w_{2,t} \). \( \Omega_{22}^{-1} \Lambda_{22} \) is the multivariate analog to \( \theta \) in the simple analytical example. Analogous to (4), \( w_{2,t} \) must be solved forward to ensure stability of the system. Sims shows that the forward solution of (18) is

\[
y_t = \Theta_1 y_{t-1} + \Theta_0 z_t + \Theta_y \sum_{s=1}^{\infty} \Theta_f^{-1} \Theta_z E_t z_{t+s} \tag{20}
\]

where \( \Theta_f = \Omega_{22}^{-1} \Lambda_{22} \), \( \Theta_z = \Omega_{22}^{-1} Q_2 \Psi \). If the structural shocks, \( z_t \), are \( i.i.d. \) and agents do not have foresight, then the last term in (20) drops out of the solution and the equilibrium has a VAR representation. By conditioning on the control and state variables, \( y_t \), an econometrician who estimates a VAR will be able to recover the structural shocks. But when agents have foresight, the equilibrium representation becomes a VARMA with the MA coefficients determined by the unstable generalized eigenvalues. Suppose the structural shocks are given by \( z_t = \epsilon_{t-q} \), and agents have foresight—at date \( t \) they observe \( \epsilon \)'s dated \( t \) and earlier. The equilibrium is

\[
y_t = \Theta_1 y_{t-1} + \Theta_0 \epsilon_{t-q} + \Theta_y [\Theta_z \epsilon_{t-q+1} + \Theta_f \Theta_z \epsilon_{t-q+2} + \cdots + \Theta_f^{q-1} \Theta_z \epsilon_t]. \tag{21}
\]

As in the univariate case, the contemporaneous shocks are discounted the heaviest, which is precisely why models with foresight are more likely to deliver non-fundamental equilibrium representations.

The extent to which foresight leads to econometric errors depends on the underlying structure of the economy and the nature of information flows. The next section examines this issue in two canonical macro models.

3. Quantitative Importance of Foresight

The information flows specification in (5) was chosen for its analytical convenience, not for its plausibility. To assess the quantitative importance of foresight, this section generalizes those flows to capture actual news processes and embeds the generalized specification in two empirically motivated DSGE models. We show how the nature of information flows affects the inference errors an econometrician can make by not modeling foresight. Quantitative importance is summarized by dynamic tax multipliers, comparing those estimated by an econometrician who fits an identified VAR to the true tax multipliers.
3.1. **Modeling Information Flows.** Rich information flows characterize the arrival and accumulation of news about tax changes, but generally fall into two periods: between initial proposal and final enactment—or rejection—of a new tax law ("inside lag") and between enactment and when the law takes effect ("outside lag"). During the inside lag, information and expectations are evolving about the likelihood and the precise form of proposed legislation. Sources of information that mark the beginning of the inside lag can be formal—a president’s State of the Union speech—or informal—a politician’s campaign pledges. And this early information may be confirmed or contravened by subsequent actions. Outside lags arise whenever there is a delay between the legislation’s passage and its implementation, as when tax changes are phased in. The two types of lags differ in important ways. During the inside lag, anticipated taxes are uncertain; news arrives regularly and induces agents to update their expectations. During the outside lag, the tax law has been adopted, no more news arrives, and agents have perfect foresight about future tax rates.

Examples clarify the nature of information flows. The Economic Recovery Tax Act of 1981, enacted in August 1981, phased in tax reductions through the beginning of 1984 to yield an outside lag of 10 quarters. In announcing his candidacy for president in November 1979, Ronald Reagan made clear that he intended to substantially lower taxes: "The key to restoring the health of the economy lies in cutting taxes" [Reagan (1979)]. News about future taxes, then, arrived throughout 1980, evolving with Reagan's prospects of winning office. An additional six months passed between President Reagan’s formal call for tax relief in February 1981 and the legislation’s enactment. The inside lag associated with this tax change is, arguably, five or more quarters, with the weights agents place on the bits of news changing over time. Taken together, the two lags imply a foresight period of about four years.

Adjustments to Social Security taxes can entail extraordinarily long lags. The National Commission on Social Security Reform was established in December 1981 to recommend solutions to the System’s short- and long-term solvency problems. Its recommendations, reported in January 1983, formed the basis for the Social Security Amendments of 1983, which were enacted in April 1983. The Amendments phased in payroll tax increases beginning in 1984 and extending to 1990. Although their inside lag may have been only a few quarters, the Amendments’ outside lag is over six years. Other changes in Social Security taxes had comparably long lags.

To model these intricacies, we generalize (5) with a specification of information flows about tax rates that is flexible enough to capture both inside and outside lags. For labor taxes, we posit

$$\hat{\tau}^L_t = \rho \hat{\tau}^L_{t-1} + \sum_{j=0}^{J} \phi_j [\sigma L L_{t-j} + \zeta \sigma K K_{t-j}]$$

(22)

---

6These labels date back to Friedman (1948), where we combine the “recognition” and “decision” lags to form inside lags and our outside lags refer to how long it takes legislation to change tax rates.

7Announcing their candidacies, both Ronald Reagan and George W. Bush made clear their intentions to cut taxes, well over a year before they took office and formally proposed tax cuts. George H. W. Bush, in contrast, pledged in his announcement speech, “I am not going to raise your taxes—period.” That was two-and-a-half years before he called for a tax increase. See http://www.4president.org for these speeches.
where \( \hat{\tau}_t^L \) is the labor tax rate, \( \xi \) permits rates to be correlated, and the \( \varepsilon \)'s are serially uncorrelated. To interpret the moving-average coefficients as weights, we impose that \( \sum_j \phi_j = 1 \).

Modeling information flows as moving average processes captures the idea that from quarter to quarter news about taxes evolves randomly. Exogenous tax processes are the best-case scenario for econometricians because identification is straightforward in the absence of foresight. This ensures that all errors arise solely from foresight.

Specification (22) embeds many of the information flows that appear in theoretical studies of foresight, including Christiano, Ilut, Motto, and Rostagno (2008), Jaimovich and Rebelo (2009), and Fujiwara, Hirose, and Shintani (2011) in the context of technology news; Ramey (2007) for government spending news; Yang (2005) and Mertens and Ravn (2011) with regard to tax news, and Schmitt-Grohé and Uribe (2008) for news about a variety of variables. These studies set \( \phi_j = 0 \) for all \( j \) except for \( \phi_q = 1 \), where \( q \) is the period of foresight. These specifications imply that once the news arrives, agents have \( q \) periods of perfect foresight about the object being modeled. This may be an adequate assumption about information flows that stem from outside lags, but they miss altogether the inside lags. Inside lags are periods when agents are learning about how the future may play out. Tax policies develop over time, from initial informal proposals to formal proposals, all the way through the legislative process. The \( \phi_j \) coefficients in (22) reflect how agents update their views about taxes during the inside lags. Values of the \( \phi_j \)'s describe how information flows differ from period to period.

3.2. Model Descriptions. We study a real business cycle model—closely related to Chari, Kehoe, and McGrattan (2008)—and a new Keynesian model—similar to those in Smets and Wouters (2003, 2007)—but add distorting tax rates on capital and labor income. These models are workhorses in the macroeconomics literature so we provide only brief descriptions here. The online appendix describes the models and our estimation strategies more thoroughly.

In the real business cycle (RBC) model, a representative agent maximizes time-separable discounted utility over consumption and leisure. The agent supplies labor and capital to a representative firm, which produces output according to a Cobb-Douglas technology. The government chooses a set of fiscal variables to satisfy the flow budget constraint, \( G_t + Z_t = \tau_t^L w_t l_t + \tau_t^R r_t^K k_{t-1} \), where \( G_t \) is government consumption, and \( Z_t \) is transfers.

Tax legislation tends to adjust labor and capital taxes simultaneously, following (22), and its analog for capital tax rates. Yang (2005) estimates the correlation between tax rates at 0.5, implying the value of \( \xi \). A “typical” tax change, analogous to those studied in VARs, moves the tax rates together.

Log-linearized government consumption policy follows the process

\[
\hat{G}_t = \rho_G \hat{G}_{t-1} + \sigma_G \varepsilon^G_t. \tag{23}
\]

Lump-sum transfers adjust to balance the government budget constraint each period.

---

8Some studies allow the news shocks, \( \varepsilon_{t-j} \), to be drawn from distinct distributions for each \( j \), and set \( \phi_j = 1 \) for each relevant \( j \) [Schmitt-Grohé and Uribe (2008), Fujiwara, Hirose, and Shintani (2011), and Mertens and Ravn (2011)]. The \( j = 0 \) shock is unanticipated, while the \( j > 0 \) shocks are anticipated given information at time \( t \).
The new Keynesian (NK) model extends the RBC model to incorporate real and nominal rigidities that have been shown to help fit macroeconomic data. It also models fiscal financing by allowing spending to adjust to stabilize government debt. The NK model adds external habit formation, differentiated labor types, a monopolistically competitive intermediate goods sector, variable capital utilization, wage and price rigidities, and a monetary authority that follows a Taylor-type rule for setting nominal interest rates. Tax policies obey (22) and government spending policies follow the process

\[ \hat{X}_t = \rho X \hat{X}_{t-1} + \gamma X \hat{S}_{t-1} + \sigma X \hat{\epsilon}_t, \quad \hat{X} \in \{ \hat{G}, \hat{Z} \} \]

(24)

where \( \hat{S}_{t-1} \equiv \frac{B_{t-1}}{Y_{t-1}} \) is the debt-output ratio and \( \gamma_X < 0 \).

We estimate the NK model using Bayesian methods and U.S. quarterly data from 1984 to 2007. To conduct simulations, we fix parameters at the mode of the posterior distributions (see table 1 in the online appendix). For the RBC model, the structural parameters are calibrated to the values used in the literature and standard deviations of the shocks are set to the values estimated in the NK model.

3.3. Information Flows and Estimation Bias. The Romers’ (2007; 2010) narrative analysis and Yang’s (2009) timeline of outside lags associated with federal tax changes reveal two critical features of information flows about taxes. First, the foresight horizon varies considerably from one piece of tax legislation to the next. Second, most tax changes entail substantial inside and outside lags. The generalized specification (22) can model these features of information flows; simple specifications like (5) cannot.

We examine the implications of four alternative information flows in the two DSGE models. The alternatives reflect the diversity of information flows that previous authors have documented. With a maximum length of tax foresight of eight quarters, the four information processes we employ appear in table 1.

Processes I and II model inside lags that differ in the intensity of information flows. In I, the flows are smooth, so news over the previous six quarters receives equal weight. Tax laws that make steady progress through the legislature and get implemented with little delay create flows like I. Process II concentrates the news on lags four through six, with small weight on recent news. Tax changes implemented with a lag of about one year, with only slight changes in details in the periods immediately before implementation, generate flows like II.

The outside lags in processes III and IV closely resemble the information flows that other authors posit [for example, Mertens and Ravn (2011), and Forni, Gambetti, and Sala (2011)]. These processes imply that agents have eight-quarter (III) or two-quarter (IV) perfect foresight about tax changes. Perfect foresight precludes any further changes in legislation, so these processes are exclusively about implementation delays or phased-in tax changes.9

9Ideally, information flows would encompass both inside and outside lags, but such flows would take us outside of a linear structure. For example, one could posit the flows for the inside lag and then, conditional on legislation having been enacted, switch to the outside lag specification, a process that is inherently nonlinear.
Table 1: Information Flow Processes. Coefficient settings in tax rule (22).

<table>
<thead>
<tr>
<th>Process</th>
<th>Lags</th>
<th>Description</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Inside 6 qtrs, smooth news</td>
<td>$\phi_j = \frac{1}{\tau}, j = 1, 2, \ldots, 6$</td>
<td>$\phi_0 = \phi_7 = \phi_8 = 0$</td>
</tr>
<tr>
<td>II</td>
<td>Inside 6 qtrs, concentrated news</td>
<td>$\phi_1 = \phi_2 = \phi_3 = 0.05, \phi_4 = 0.25$</td>
<td>$\phi_5 = \phi_6 = 0.3, \phi_0 = \phi_7 = \phi_8 = 0$</td>
</tr>
<tr>
<td>III</td>
<td>Outside 8-qtr phase-in</td>
<td>$\phi_j = 0, j \neq 8$</td>
<td>$\phi_8 = 1$</td>
</tr>
<tr>
<td>IV</td>
<td>Outside 2-qtr phase-in</td>
<td>$\phi_j = 0, j \neq 2$</td>
<td>$\phi_2 = 1$</td>
</tr>
</tbody>
</table>

Table 2 summarizes the actual and estimated output multipliers associated with a typical tax change in the RBC and NK models. In this exercise, the agent knows the information process and observes the actual $\varepsilon_t$'s. The econometrician, on the other hand, bases inference on a set of observable variables. We construct the innovations representation based on the econometrician’s conditioning set and use the Kalman filter to back out the econometrician’s inferences about the responses of output and taxes to a shock to the tax rate. For the RBC model, the econometrician conditions on the labor tax rate, income tax revenue, output, and investment; the conditioning set for the NK model adds government consumption, private consumption, labor, government debt, inflation, and the nominal interest rate. Thus, the estimated VAR contains several “forward-looking” variables. As a robustness check, we examined many combinations of alternative conditioning variables and found results that are consistent with those in table 2. We report biases as estimated less actual multipliers and biases as a percentage of the actual multipliers. In the absence of foresight, the bias is always zero.

Several general findings emerge from the table. Biases can be very large—hundreds of percent—and can change sign over time across both models. In both models, the biggest errors arise from outside lags—information processes III and IV—which are the information flows most frequently posited in work on foresight. Inside lags with moving-average terms—processes I and II—produce smaller, though still sizeable errors. Information process III, in which agents have two years of perfect foresight about tax rates, generates the largest inference errors in both models. It also confounds dynamics: the econometrician estimates that the strongest effect is contemporaneous, while the largest impact actually occurs two or three years later, depending on the model.

In the RBC model, actual multipliers change sign—positive in the foresight period and negative later—but estimated multipliers are uniformly negative. Frictions in the NK model propagate errors, making short/long-run distinctions less pronounced.\(^\text{10}\) In the frictionless RBC model, biases dissipate over time.

A consistent finding across the two models is that for horizons of eight quarters and beyond, the econometrician underestimates the multiplier. The lone exception being the new Keynesian model under information process I. The discounting of the tax innovations

\(^{10}\)This echoes Leeper and Walker’s (2011) results for foresight about technology.
that appears in (4) and (20) explains this result. An agent with \( q \) quarters of foresight discounts the innovations so that the \( \varepsilon_{t-q} \) shock receives little discount relative to shocks dated \( t \) through \( t - q - 1 \). As in the analytical model, this perverse discounting occurs because \( \varepsilon_{t-q} \) informs about the contemporaneous tax rate, \( \tau_t \), while shocks dated \( t \) through \( t - q - 1 \) inform about future tax rates. An econometrician, who does not observe the true innovations, applies the conventional discounting to the innovations in her information set, as in (11). This makes the econometrician’s impulse response functions die out faster than the true impulse response functions, yielding the underestimates.

<table>
<thead>
<tr>
<th>Info Process</th>
<th>0 qtr</th>
<th>4 qtrs</th>
<th>8 qtrs</th>
<th>12 qtrs</th>
<th>20 qtrs</th>
<th>peak (qtr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I actual</td>
<td>0.19</td>
<td>-1.14</td>
<td>-1.48</td>
<td>-1.11</td>
<td>-0.65</td>
<td>-1.71 (6)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.31</td>
<td>-1.35</td>
<td>-1.27</td>
<td>-0.97</td>
<td>-0.59</td>
<td>-1.57 (5)</td>
</tr>
<tr>
<td>bias</td>
<td>-0.50</td>
<td>-0.21</td>
<td>0.20</td>
<td>0.14</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-263%</td>
<td>-19%</td>
<td>14%</td>
<td>12%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>II actual</td>
<td>0.15</td>
<td>-0.54</td>
<td>-1.40</td>
<td>-1.05</td>
<td>-0.61</td>
<td>-1.62 (6)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.56</td>
<td>-1.46</td>
<td>-1.19</td>
<td>-0.91</td>
<td>-0.55</td>
<td>-1.48 (2)</td>
</tr>
<tr>
<td>bias</td>
<td>-0.71</td>
<td>-0.92</td>
<td>0.21</td>
<td>0.14</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-473%</td>
<td>-169%</td>
<td>15%</td>
<td>13%</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>III actual</td>
<td>0.09</td>
<td>0.16</td>
<td>-1.51</td>
<td>-1.12</td>
<td>-0.64</td>
<td>-1.51 (8)</td>
</tr>
<tr>
<td>estimated</td>
<td>-1.44</td>
<td>-1.09</td>
<td>-0.82</td>
<td>-0.64</td>
<td>-0.39</td>
<td>-1.44 (0)</td>
</tr>
<tr>
<td>bias</td>
<td>-1.54</td>
<td>-1.24</td>
<td>0.69</td>
<td>0.49</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-1641%</td>
<td>-784%</td>
<td>46%</td>
<td>43%</td>
<td>39%</td>
<td></td>
</tr>
<tr>
<td>IV actual</td>
<td>0.16</td>
<td>-1.34</td>
<td>-1.00</td>
<td>-0.76</td>
<td>-0.45</td>
<td>-1.56 (2)</td>
</tr>
<tr>
<td>estimated</td>
<td>-1.41</td>
<td>-1.06</td>
<td>-0.81</td>
<td>-0.62</td>
<td>-0.38</td>
<td>-1.41 (0)</td>
</tr>
<tr>
<td>bias</td>
<td>-1.57</td>
<td>0.28</td>
<td>0.20</td>
<td>0.14</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-962%</td>
<td>21%</td>
<td>20%</td>
<td>18%</td>
<td>16%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Info Process</th>
<th>0 qtr</th>
<th>4 qtrs</th>
<th>8 qtrs</th>
<th>12 qtrs</th>
<th>20 qtrs</th>
<th>peak (qtr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I actual</td>
<td>-0.08</td>
<td>-0.36</td>
<td>-0.48</td>
<td>-0.43</td>
<td>-0.24</td>
<td>-0.48 (8)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.07</td>
<td>-0.44</td>
<td>-0.57</td>
<td>-0.51</td>
<td>-0.28</td>
<td>-0.57 (8)</td>
</tr>
<tr>
<td>bias</td>
<td>0.01</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.08</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>11%</td>
<td>-24%</td>
<td>-20%</td>
<td>-18%</td>
<td>-18%</td>
<td></td>
</tr>
<tr>
<td>II actual</td>
<td>-0.06</td>
<td>-0.27</td>
<td>-0.43</td>
<td>-0.40</td>
<td>-0.23</td>
<td>-0.43 (9)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.09</td>
<td>-0.37</td>
<td>-0.42</td>
<td>-0.37</td>
<td>-0.19</td>
<td>-0.42 (7)</td>
</tr>
<tr>
<td>bias</td>
<td>-0.03</td>
<td>-0.10</td>
<td>0.00</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-51%</td>
<td>-37%</td>
<td>1%</td>
<td>9%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>III actual</td>
<td>-0.03</td>
<td>-0.12</td>
<td>-0.32</td>
<td>-0.37</td>
<td>-0.26</td>
<td>-0.37 (12)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.14</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.01</td>
<td>-0.14 (0)</td>
</tr>
<tr>
<td>bias</td>
<td>-0.11</td>
<td>0.01</td>
<td>0.24</td>
<td>0.32</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-340%</td>
<td>13%</td>
<td>76%</td>
<td>85%</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>IV actual</td>
<td>-0.06</td>
<td>-0.30</td>
<td>-0.33</td>
<td>-0.28</td>
<td>-0.14</td>
<td>-0.33 (7)</td>
</tr>
<tr>
<td>estimated</td>
<td>-0.15</td>
<td>-0.24</td>
<td>-0.26</td>
<td>-0.22</td>
<td>-0.11</td>
<td>-0.26 (7)</td>
</tr>
<tr>
<td>bias</td>
<td>-0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>% bias</td>
<td>-128%</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Output Multipliers for a Labor Tax Change, Correlated with a Capital Tax Change. Multipliers are output responses scaled by the peak response of revenues, converted to dollars, as in Blanchard and Perotti (2002). Agent knows the information process and observes the actual \( \epsilon_t \)’s. Econometrician bases inference on a set of observable variables, as described in text. Biases equal estimated less actual multipliers.
These findings establish two key points. First, failure to model fiscal foresight can produce quantitatively important errors of inference in the canonical models used for macroeconomic policy analysis. Second, the precise nature of information flows about news matters for the pattern of inference errors. Getting the information flows “right” poses a substantial challenge to DSGE modelers. We turn now to empirical approaches designed to address the errors associated with foresight.

4. Attacking the Problem

This section offers a unification of the empirical approaches that have been introduced to deal with the econometric problems associated with foresight. We show how seemingly diverse approaches [e.g., the narrative approach of Ramey (2007) and Romer and Romer (2010) and the dynamic stochastic general equilibrium approach of Schmitt-Grohé and Uribe (2008)], are attempts at solving similar problems caused by foresight. We provide a brief discussion of the pros and cons of each approach, while offering detailed calculations in support of the discussion in an online appendix.

4.1. Non-uniqueness of Moving Average Representations. Moving average representations are not unique for two distinct reasons that Hansen and Sargent (1991a) emphasize. Understanding the reasons for non-uniqueness will provide a useful way to characterization solutions to the problems generated by foresight. Consider the Wold representation for the $n \times 1$ vector stochastic process $x_t$

$$x_t = \sum_{j=0}^{\infty} H_j^* \epsilon^*_{t-j}$$ (25)

where $\sum_{j=0}^{\infty} \text{tr } H_j^* H_j^t < \infty$ and $\epsilon^*_t$ is an $n$-dimensional white noise process defined as the innovation in predicting $x_t$ linearly from its semi-infinite past ($\epsilon^*_t \equiv x_t - P[x_t|x^{t-1}]$).

There are two transformations that are observationally equivalent to (25). The first comes from multiplying by a nonsingular matrix $U$,

$$x_t = \sum_{j=0}^{\infty} (H_j^* U^{-1})(U \epsilon^*_{t-j})$$ (26)

where the innovation is now defined as $U \epsilon^*_t$ and $H_j^* U^{-1}$ represents the altered impulse responses. If $U$ is nonsingular, then the new innovations process spans the same space as $x_t$ and the information content of $U \epsilon^*_t$ is identical to that $\epsilon^*_t$. This is the type of non-uniqueness that Sims (1980) describes. Researchers confront this non-uniqueness with different orthogonalization schemes that rotate the covariance matrix through recursive orderings [Sims (1980)], short-run restrictions [Bernanke (1986), Sims (1986)], long-run restrictions [Blanchard and Quah (1989)], a combination of short and long-run restrictions [Gali (1999)], or sign restrictions [Faust (1998), Canova (2000), Uhlig (2005)].
A second type of non-uniqueness that is observationally equivalent to (25) is generated by foresight and is described by the non-fundamental representation

\[ x_t = \sum_{j=0}^{\infty} H_j \epsilon_{t-j} \]  

(27)

where now \( \{\epsilon_{t-j}\}_{j=0}^{\infty} \) spans a larger space than \( \{x_{t-j}\}_{j=0}^{\infty} \), and \( H(L) \) satisfies

\[ H^*(z) E \epsilon_t^* \epsilon_t^\prime H^*(z^{-1})' = H(z) E \epsilon_t \epsilon_t' H(z^{-1})' \]

Under the typical assumption that agents observe the structural shocks \( \epsilon_t \) directly while the econometrician observes only \( x_t \), models with sufficient foresight belong to this class of non-fundamental representations. The covariance generating functions of \( H(L) \epsilon_t \) and \( H^*(L) \epsilon_t^* \) are identical, but only \( H^*(L) \) possesses an invertible representation in \( x_t \). Let \( A(L) = H^*(L)^{-1} \). The typical VAR methodology delivers

\[ x_t = A_0^{-1} [A_1 x_{t-1} + A_2 x_{t-2} + \cdots + \epsilon_t^*] \]  

(28)

Identifying \( A_0^{-1} \) in the usual way recovers the shocks \( \epsilon_t^* \), but not the structural shocks, \( \epsilon_t \), that agents observe. Thus the econometrician will condition on a smaller information set than the agent’s.

Hansen and Sargent’s non-uniqueness point sends a clear message: to identify structural shocks in a vector autoregression, both types of non-uniqueness must be confronted. Confronting the non-uniqueness in (26) does not solve the non-uniqueness of representation (27), and vice versa. Although a large literature focuses on the non-uniqueness associated with (26), identifying (27) requires the econometrician to condition on the same information set as the agents they are modeling.

Casting the problem as resolving the two distinct forms of non-uniqueness sheds light on approaches that appear in the empirical macro literature. One line of attack estimates conventional VARs, identified in a variety of creative ways to isolate anticipated effects, and then examines the impacts of foresight [Sims (1988), Blanchard and Perotti (2002), Yang (2007), Mountford and Uhlig (2009), Beaudry and Portier (2006), Fisher and Peters (2009), Barsky and Sims (2011)]. For example, Beaudry and Portier (2006) and Fisher and Peters (2009) condition on stock prices to capture news about expected changes in technology and government spending, respectively. Barsky and Sims (2011) identify news about productivity as the shock that is orthogonal to current utilization-adjusted productivity that best explains future variations in adjusted productivity. A second line of attack rejects VAR identification schemes, arguing that VARs cannot adequately measure the impacts of foreseen changes in fiscal policy, and takes a narrative approach to identification that brings fresh data to bear on the problem [Ramey and Shapiro (1998), Edelberg, Eichenbaum, and Fisher (1999), Burnside, Eichenbaum, and Fisher (2004), Ramey (2007), Romer and Romer (2010)]. A third approach uses standard methods, such as An and Schorfheide (2007), to estimate a model with foresight. To pursue this approach, Schmitt-Grohé and Uribe (2008) make very particular assumptions about the information flows that give rise to foresight about technology and government spending. The tradeoff is that the modeler must be explicit about the role of information in the economy.
4.2. Identifying Information Flows. A unifying theme underlying each approach, though, is the need to specify an information structure that is consistent with agents that have foresight. Although most papers do not explicitly cast the problem as resolving the non-uniqueness of (27) by specifying information flows, this is precisely what each paper aims to do in its own way. We have seen in table 2 that different assumptions about information flows can dramatically alter the dynamics of the model and the associated econometric inferences. Careful modeling of the information available to agents is of first-order importance.

4.2.1. Conditioning on Asset Prices. If asset markets are efficient, the information contained in asset prices should coincide with all available information to agents. Conditioning on asset prices in VARs should help align the information set of the econometrician and agent. With respect to fiscal foresight, there exists an asset class that is particularly useful for isolating news about future tax shocks. In the United States, municipal bonds are exempt from federal taxes.\(^{11}\) The differential treatment of municipal and treasury bonds has useful implications for identifying news about tax changes. If \(Y_t^M\) is the yield on a municipal bond at \(t\) and \(Y_t^T\) is the yield on a taxable bond, and assuming the bonds have the same term to maturity, callability, market risk, credit risk, and so forth, then an “implicit tax rate” is given by \(\tau_t^I = 1 - \frac{Y_t^M}{Y_t^T}\). This is the tax rate at which the investor is indifferent between tax-exempt and taxable bonds. If participants in the municipal bond market are forward looking, the implicit tax rate should predict subsequent movements in individual tax rates. For example, if investors expect individual tax rates to rise (fall), they will demand higher (lower) yields on taxable bonds until they are indifferent between taxable and nontaxable bonds.

In Leeper, Walker, and Yang (2011) and the online appendix, we show that implicit tax rates are Granger-causally prior relative to the information sets in the fiscal VAR systems that Blanchard and Perotti (2002) and Mountford and Uhlig (2009) estimate. This result is consistent with the ability of the municipal bond market to forecast changes in fiscal policy [Poterba (1989), Fortune (1996), ?]. Employing exactly the identification schemes and data sets of Blanchard-Perotti and Mountford-Uhlig, we ask how augmenting the econometrician’s information set with a direct measure of tax news affects inferences. As shown in Leeper, Walker, and Yang (2011) and the online appendix, for Blanchard-Perotti, results change dramatically: anticipated tax increases raise output substantially for about three years before output begins to decline. Differences also emerge for Mountford-Uhlig. Investment multipliers, which Mountford-Uhlig estimated to be zero, become significantly positive.


For example, Mertens and Ravn (2011) augment a VAR with Romer and Romer’s (2010) anticipated tax liabilities series, which they treat as strictly exogenous in the VAR. Mertens and Ravn append to each equation of the VAR a distributed lag of \(q\) periods in future

\(^{11}\)Depending upon the type of bond, municipal bonds can also be exempt from the Alternative Minimum Tax, state, and local taxes. See Ang, Bhansali, and Xing (2010) for a thorough description of the municipal bond market.
tax liabilities. They estimate that an anticipated tax increase induces a boom in output whose amplitude and duration increase with the period of foresight $q$. One downside to this approach, in contrast to introducing the muni-treasury spreads directly into the VAR, is that one must specify a priori the period of foresight and maintain that anticipated taxes are exogenous—assumptions that are critical to the quantitative effects they obtain. Nonetheless, the qualitative effects of Mertens and Ravn (2011) are similar to the results one would get by conditioning on implicit tax rates (see online appendix figure 4 panel A). This should not be surprising because despite their different approaches to solving the non-uniqueness of (27), the narrative approach and the asset pricing approach share a common economic explanation for their findings. Anticipated tax changes generate wealth effects that kick in immediately—upon arrival of the news—but the substitution effects, which operate on critical economic margins, do not affect behavior until the tax rates have changed. In a conventional model, expected tax increases reduce wealth, which induces agents to work harder, increasing employment and output immediately.

4.2.3. Estimation of DSGE Model. A third approach uses standard methods, such as An and Schorfheide (2007), to estimate a model with foresight. By specifying the entire structure of the economy, including the information flows process, one is free to estimate the moving average representation directly and does not have to worry about the invertibility of the moving average representation. However, the tradeoff is that one must make very particular assumptions about the information flows that give rise to foresight about technology and government spending, taxes, etc. Surprisingly, despite the centrality of information structures to the burgeoning news literature—whether it be news about technology or news about fiscal or monetary actions—there has been essentially no exploration of alternative, equally plausible, assumptions about how information about critical economic variables flows to agents [Leeper and Walker (2011)].

5. Concluding Remarks

We have shown how foresight introduces econometric difficulties that complicate the interpretation of conventional econometric analyses. Foresight, of any type, can introduce non-fundamental moving average terms into the linear equilibrium process, changing the mapping between the true news that agents observe and the “shocks” that the econometrician identifies. Many of the econometric techniques in macroeconomists’ toolboxes can be distorted by empirical methods that do not adequately estimate the non-invertible moving average components of equilibrium time series. Section 3 demonstrates that failing to model foresight can produce quantitatively important inference errors in data generated by models now in wide use for macro policy analysis. Section 4 employs municipal bond spreads to capture information flows about anticipated changes in tax rates. Incorporating this spread into VARs and imposing well-known identification schemes can drastically alter conclusions about the dynamic effects of anticipated tax changes.

Estimating the impacts of foresight requires either modeling the information flows about future economic fundamentals or finding direct and interpretable measures of news. In the former camp are efforts to estimate DSGE models with news and the closely related approach that “flips” the roots of the invertible process to obtain the non-invertible representation that
FISCAL FORESIGHT AND INFORMATION FLOWS

forecasts resolve the econometric problems that foresight presents by making strong assumptions about information flows. Of course, the solutions are conditional on the specified information flows, aspects of the economic structure about which economists rarely have well developed prior beliefs or direct empirical evidence.12

Some authors, rather than seeking direct measures of news, rely on adding variables to try to align the econometrician’s and the agents’ information sets. Forni and Gambetti (2010b) estimate large empirical models that boil all relevant information down to a few critical factors and test whether a model’s information content is “sufficient” to ensure fundamentalness. Their sufficient condition, however, will never be satisfied by the small-to-medium-sized VARs that have been heavily used to extract economically interpretable shocks. Because the estimated factors do not have clear economic interpretations, it is impossible to discern precisely what information is being tested for.13 In contrast, our approach to tax foresight or Ramey’s (2011) method for measuring news about government defense spending focus narrowly on a particular, economically unambiguous, type of news.

Foresight poses a challenging mix of structural and measurement problems. Hypothesized information flows that are uninformed by observations and information sets that are unrestricted by theory are unlikely to resolve the foresight problem. Answers lie in blending theory with measurement.

References


12 That the assumptions about information flows matter to inferences from estimated DSGE models is shown in two versions of the same paper that differ in information flow specifications and yield very different inferences [Schmitt-Grohé and Uribe (2008) and Schmitt-Grohé and Uribe (2010)].
13 Tests for invertibility harken back to the 0−1 treatment of invertibility from which this paper advocates moving beyond.


