Macroeconomic Models for Monetary Policy: A Critical Review from a Finance Perspective

Winston W. Dou,1 Andrew W. Lo,2,3 Ameya Muley,4 and Harald Uhlig5

1Department of Finance, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: wdou@wharton.upenn.edu
2Sloan School of Management, Laboratory for Financial Engineering, and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA; email: alo-admin@mit.edu
3Santa Fe Institute, Santa Fe, New Mexico 87501, USA
4Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA; email: ameyamuley@gmail.com
5Department of Economics, University of Chicago, Chicago, Illinois 60637, USA; email: huhlig@uchicago.edu

Keywords
macrofinancial models, dynamic stochastic general equilibrium models, macroprudential policy, systemic risk, monetary policy analysis

Abstract
We provide a critical review of macroeconomic models used for monetary policy at central banks from a finance perspective. We review the history of monetary policy modeling, survey the core monetary models used by major central banks, and construct an illustrative model for those readers who are unfamiliar with the literature. Within this framework, we highlight several important limitations of current models and methods, including the fact that local-linearization approximations omit important nonlinear dynamics, yielding biased impulse-response analysis and parameter estimates. We also propose new features for the next generation of macrofinancial policy models, including a substantial role for the financial sector, the government balance sheet, and unconventional monetary policies; heterogeneity, reallocation, and redistribution effects;
the macroeconomic impact of large nonlinear risk premium dynamics; time-varying uncertainty; financial sector and systemic risks; imperfect product market and markups; and further advances in solution, estimation, and evaluation methods for dynamic quantitative structural models.

1. INTRODUCTION

The Financial Crisis of 2008 and the Great Recession that followed revealed serious gaps in commonly used approaches to define, measure, and manage financial sector activities that pose risks to the macroeconomy as a whole.

One emerging narrative is that macroeconomic models commonly employed at the institutions evaluating monetary policy lack the analytical specificity to account for important financial sector influences on the aggregate economy. A new generation of enhanced models and advanced empirical and quantitative methodologies are needed by policy makers and must be developed by researchers to better study the impact of shocks that are initially large or build endogenously over time.

This article presents a review of these macroeconomic models and their empirical methods. Through this review, we hope to clarify the most important challenges of existing macroeconomic models for monetary policy analysis and summarize some recent advances in new modeling and quantitative techniques. The primary goal of this review is to provide insight, guidance, and motivation for the next generation of young scholars—especially those at the intersection of macroeconomics and financial economics—to develop more effective macroeconomic models to inform policy decisions.

There has been a remarkable evolution of macroeconomic models used for monetary policy at major central banks around the world, in aspects such as model formulation, solution methods, estimation approaches, and the communication of results between central banks. Central banks have developed many different classes and variants of macroeconomic models in the hope of producing a reliable and comprehensive analysis of monetary policy. Early models included quantitative macroeconomic models,1 reduced-form statistical models, structural vector autoregressive models, and large-scale macroeconometric models, a hybrid form combining the long-run structural relationships implied by a partial equilibrium treatment of theory (e.g., the decision rule for aggregate consumption) and reduced-form short-run relationships employing error-correcting equations.

Over the past 20 years, there have been significant advances in the specification and estimation for New Keynesian dynamic stochastic general equilibrium (DSGE) models. Significant progress has been made to advance policy-making models from the older static and qualitative New Keynesian style of modeling to the New Keynesian DSGE paradigm. A New Keynesian DSGE model is designed to capture real-world data within a tightly structured and self-consistent macroeconomic model. It has explicitly theoretical foundations, allowing it to circumvent the Sims critique (see Sims 1980) and the Lucas critique (see Lucas 1976); therefore, it can provide more reliable monetary policy analysis than earlier models.

A consensus baseline New Keynesian DSGE model has emerged, one that is heavily influenced by estimated impulse-response functions based on structural vector autoregression (SVAR) models. In particular, Christiano, Eichenbaum & Evans (2005) have shown that a baseline New

---

1 Examples include the Wharton econometric model and the Brookings model.
Keynesian DSGE model can successfully account for the effects of a monetary policy shock with nominal and real rigidities. Similarly, Smets & Wouters (2003, 2007) show that a baseline New Keynesian DSGE model can track and forecast time series as well as, if not better than, a Bayesian vector autoregression (BVAR) model. New Keynesian DSGE models have been developed at many central banks, becoming a key part of many of their core models.\(^2\) Sbordone et al. (2010) have emphasized that an advantage of New Keynesian DSGE models is that they share core assumptions about the behavior of agents, making them scalable to relevant details to address the policy question at hand. For example, Smets & Wouters (2007) introduced wage stickiness and investment frictions into their model; Gertler, Sala & Trigari (2008) and Blanchard & Galí (2010) incorporated labor market search and wage bargaining; and Bernanke, Gertler & Gilchrist (1999), Chari, Christiano & Eichenbaum (1995), and Christiano, Motto & Rostagno (2008) studied the interaction between the financial sector and macroeconomic activity. Interestingly, DSGE models with richer structures have been included in core models of several central banks.\(^3\)

However, the devastating aftermath of the Financial Crisis of 2008 and the subsequent Great Recession has prompted another rethink of monetary and central banking policies, which are now facing many new challenges. Most macroeconomists and many policy makers and regulators have called for a new generation of DSGE models. The first and foremost critique of the current state of standard New Keynesian DSGE models is that they lack an appropriate financial sector with a realistic interbank market and that, as a result, they fail to fully account for an important source of aggregate fluctuations, such as systemic risk from the financial system. Second, the linkage between the endogenous risk premium and macroeconomic activity is central to our understanding of the transmission mechanisms of monetary policy, especially during financially stressed periods. In models that lack a coherent endogenous risk premium, policy experiments become unreliable in stressed periods, and the model cannot provide a consistent framework for conducting experimental stress tests regarding financial stability or macroprudential policy. Third, heterogeneity among the players in the economy is essential to our understanding of inefficient allocations and flows between agents. These inefficiencies may have important effects on the equilibrium state of the economy.

Policy makers will need a new generation of models with a unified and coherent framework for both conventional and unconventional monetary policies. For example, at the onset of the financial crisis, the zero lower bound for short-term interest rates went from a remote possibility to reality with frightening speed. This led central banks to quickly develop unconventional measures to provide economic stimulus, including credit easing, quantitative easing, and extraordinary forward guidance. These unconventional measures require a proper platform to be analyzed. Furthermore, these measures have blurred the boundary between monetary policy and fiscal policy. Through these policies, central banks gave preference to some debtors over others (e.g., industrial companies, mortgage banks, governments), and some sectors over others (e.g., export versus domestic). In turn, the distributional effects of monetary policy were much stronger than in normal times; hence, these measures are sometimes referred to as quasi-fiscal policy. As Sims (1980)

\(^2\) The Bank of Canada, the Bank of England, the Central Bank of Chile, the European Central Bank, Norges Bank, Sveriges Riksbank, and the US Federal Reserve have all incorporated New Keynesian DSGE models into their core models.

\(^3\) Nevertheless, these New Keynesian DSGE models are by no means perfect. The models’ microeconomic foundations (for example, adjustment costs of all kinds) are basically reverse-engineered to fit the important patterns of macroeconomic variables. Furthermore, in their simplest forms, the two core behavioral equations of the model (i.e., the Keynesian Phillips curve and IS curve) are still empirically questionable.
emphasized, a reliable monetary policy experiment cannot ignore the effect of ongoing fiscal policy. In order to implement unconventional measures during the crisis, central banks put much more risk onto government balance sheets than ever before, which had the potential to lead to substantial losses. Thus, the government balance sheets in these models should be forward looking, and their risk characteristics are critical to the success of the model.

Methodological and empirical challenges have arisen along the way. First, advanced nonlinear solution methods and estimation approaches are necessary if one wishes to guarantee that key nonlinear dynamics in the financial market and the macroeconomy are eventually captured in quantitative analysis. Second, data availability and risk measurement are always a central challenge in macroeconomic modeling, but especially so in the wake of the global financial crisis and the subsequent global economic recession. Brunnermeier, Gorton & Krishnamurthy (2012) pointed out that our current measurement systems are outmoded, leaving regulators, academics, and risk managers in a dangerous position. Assessing systemic risk requires viewing data on the financial sector through the lens of a macroeconomic model. However, macroeconomics in particular frames questions and builds models based on available data, and so far we have lacked the data to construct macrofinancial models. New infrastructure for detailed microlevel financial data collection is necessary and critical for further risk measurement development and model construction. In fact, the Office of Financial Research at the US Department of the Treasury already has this mandate, and the first steps toward a new, comprehensive financial data collection system are under way.

These are the issues we hope to address in this review. In Section 2 we present a brief summary of the goals and mechanisms of central banking monetary policy, a history of macroeconomic policy models, and some motivation for the most popular framework today: the DSGE model. In Section 3 we consider a series of critiques of the current DSGE framework as well as suggestions for future directions for developing the next generation of DSGE policy models. We conclude in Section 4, and to motivate readers to take a more active interest in practical applications of policy modeling, in the Supplemental Appendices we present a survey of the core models employed by the US Federal Reserve (Fed), the European Central Bank (ECB), the Bank of England, and the Bank of Canada. Most of these models are well documented, and young scholars are encouraged to develop improvements that could have an enormous impact on macroeconomic policy and society.

2. CENTRAL BANKS, MONETARY POLICY, AND MODELS

According to a 1977 amendment to the Federal Reserve Act, the Fed's monetary policy has three basic objectives. These are to promote a "maximum" sustainable output and employment, a moderate long-term interest rate, and a level of "stable" prices. These three basic goals of monetary policy are shared by most major central banks. For example, the Treaty on the Functioning of the European Union also promotes the primary objective of supporting stable prices.

---

4To accelerate progress in macrofinancial policy modeling, we present a fully specified canonical example of the DSGE model with financial sectors in a companion article (Dou et al. 2019) that readers can work with immediately (an open-source software implementation is provided at the authors’ web pages). The model in the companion article is solved globally. We hope that the contribution of our code and global solution method in the companion article, together with this review, may be of general interest to a broader group of researchers in the macrofinancial and monetary economics community.

5The terms “price stability” and “inflation stability” are often used synonymously, and we do the same in this review.
Price stability is an economic environment that avoids both prolonged inflation and deflation. In such an environment, households and firms can make financial decisions without worrying about where prices are headed.\(^6\)

It has been argued that a monetary policy directed at maintaining aggregate price level stability will lessen both the incidence and severity of financial instability, most famously in the Schwartz hypothesis (e.g., Schwartz 1988, 1995). Obviously, price stability does not guarantee financial stability, and the Financial Crisis of 2008 provides an excellent case in point. Some researchers (e.g., Borio & Lowe 2002; Brunnermeier, Gorton & Krishnamurthy 2012; Brunnermeier & Sannikov 2014; Duarte & Eisenbach 2013) have even argued that periods of tranquility can aid the buildup of financial instability.

There has been a long debate on whether the central bank is the natural guarantor of the stability of the financial system. In this review, we emphasize the natural responsibility of the central bank in this matter for the following reasons. First, the central bank is the only provider of the legal means of payment; therefore, it is the only provider of immediate liquidity during a financial crisis. Second, a natural role of the central bank is to ensure the smooth functioning of the national payment system. As such, it is centrally positioned to monitor and combat systemic risk, defined here as the risk of the serious impairment of a large or critical part of the financial system. Third, the financial system is the transmission mechanism through which monetary policy has its effect on the real economy. The status of the financial system is critical for the central bank to have any desirable impact and achieve its monetary objectives. For this reason alone, central banks have a natural interest in maintaining a sound financial system. Finally, financial stability may play an important role in guaranteeing price stability, already a basic role of the central bank. A discussion of the trade-off between financial conditions and financial stability is presented by Adrian & Liang (2014), among others.

It is also one of a central bank’s main responsibilities to maintain a sound central bank balance sheet. Central bank balance sheets have proved crucial in designing and understanding policies pursued in the wake of the Financial Crisis of 2008. In particular, large-scale asset purchase programs became the primary tools in efforts to prevent any renewal of the financial meltdown as the effective zero lower bound for interest rates was reached. With short-term interest rates near zero, and the effectiveness of conventional monetary policies constrained as a result of a liquidity trap, these policies sought to provide additional monetary stimulus by lowering the long-term interest rate on government bonds. A loss of confidence in banks and in many financial products in the advanced economies disrupted global financial markets. This occurred in large part because the normal operations of financial markets became impaired, blocking the transmission of lower policy rates to the real economy. Central banks countered this reaction by buying unconventional assets on a large scale. They started with short-term lending, or by buying short-term assets, but progressively moved toward buying long-term assets. At present, the aggregate size of central bank balance sheets in advanced countries is nearly $8 trillion, the equivalent of more than 20% of GDP. In some cases, balance sheets are still growing.

Whether the risks that may be inherent in these large balance sheets matter, and if so, how much, merits detailed attention going forward. A country is surely better off if its central bank has the full financial strength needed to carry out its functions.

\(^6\)The ECB’s Governing Council announced a quantitative definition of price stability in 1998: A year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below 2% (Gerdesmeier 2009).
2.1. The Mechanisms and Tools of Monetary Policy

The scope of monetary policy is limited in terms of what variables the central banks can directly control, to what extent, and for how long the impact of monetary policy will last. The process of how the shock in monetary policy leads to changes in aggregate economic variables—including inflation, output, employment, consumption, and investment—is known as the monetary policy transmission mechanism. In order to analyze the monetary policy transmission mechanism, an admittedly superficial, but nonetheless helpful, approach is to think about the particular channels through which monetary policies operate and affect the real economy. On one hand, this approach is superficial since ultimately it is the general equilibrium dynamic interdependence of a solution for all the variables that determines the outcome. Examining individual “channels” can be misleading for that reason. Furthermore, examining channels is unlikely to be of much help for a precise, quantitative analysis. On the other hand, this approach is helpful because thinking about the complexities of general equilibrium can be difficult, and thinking about channels can be intuitive and instructive. Furthermore, it most easily expresses conventionally held views about how monetary policy affects the economy.

The most common channels arising in discussions of monetary policy are the interest rate channel, the inflation expectations channel, the balance sheet channel, the bank credit channel or bank lending channel, the exchange rate channel, and the asset price channel. Model builders often seek to construct models that incorporate these channels, or else provide theories with a more contrarian view. Here we only seek to describe the conventional perspective on these channels in, admittedly, rather loose and imprecise language, and not to raise questions about the validity of these conventional perspectives per se. One should not, of course, confuse models constructed to exhibit these channels with evidence that these channels are actually present. It is conceivable that considerable progress can be made by breaking through the apparent circularity of such discussions. Figure 1 presents an overview, together with a list of commonly used shocks as sources of fluctuations and a stylized description of the economy overall.

Let us first consider the interest rate channel, in which lower nominal short-term interest rates lead to lower real interest rates because prices are sticky. These lower interest rates, in turn, promote investment and consumption, but discourage savings. Conversely, higher interest rates stimulate savings and lower consumption and investment in the short run. Matters are more complicated when examined in detail, of course. There is no such thing as a single interest rate. A change in the official short-run interest rates directly affects money market interest rates, but it only indirectly affects lending and deposit rates, which are set by commercial banks to their debtors and depositors, respectively. While the central bank can control short-term interest rates, the real economy is affected mainly by the medium- and long-term deposit and lending rates charged by these commercial banks to their customers. These rates depend not only on the interest rate set by the monetary authority but also on a number of other determinants, such as inflation expectations and the risk premium of other channels. The benchmark consensus stipulates that monetary policy has almost no influence on the long-run aggregate supply. Thus, in the short to medium run, monetary policy can influence only the difference between the actual level of economic activity and the one that is sustainable over the long run, the potential output. This difference is called the output gap.

We next consider the inflation expectations channel in which looser monetary policy will result in more inflation down the road, which in turn determines economic choices in the present. At given nominal interest rates, increases in expected inflation lower the real rate, resulting in higher consumption and lower savings. The resulting inflation dynamics will reinforce inflation expectations. Higher demand for consumption or investment goods will put pressure on their marginal
cost of production. Faced with an increase in production costs, some firms might decide to reduce their profit margins, but plausibly, many firms will gradually transfer these costs to the final price, which will eventually generate realized inflation from inflation expectations. Inflation expectations are heavily affected by the perceptions of economic agents regarding the central bank’s commitment toward achieving its primary objectives. Anchoring inflation expectations can be one of the most powerful and efficient channels of monetary policy transmission, provided that it is transparent and its actions are regarded as credible. This is easier said than done. As pointed out by Blanchard (2009, p. 222), “although we very much want to believe that monetary policy can anchor inflation expectations, I am not sure we actually understand whether and how it can actually achieve it.”

The balance sheet channel is deeply associated with the external finance premium, which is defined as the wedge between the cost of capital internally available to firms and the cost of raising capital externally by issuing equity or borrowing from corporate debt markets. In the balance sheet channel, an interest rate increase worsens a firm’s balance sheet and raises its external finance premium because it raises the firm’s debt burden through higher interest payments on floating rate debt, lowers the value of the firm’s business via the overall reduction in demand, and decreases the value of the firm’s collateral through decreased asset prices. In turn, an increase in the external finance premium makes firms more reluctant to invest and expand, resulting in a decrease in aggregate economic activity. The balance sheet channel is potentially dangerous, as it could amplify and propagate small fluctuations via a pecuniary externality or an adverse feedback

**Figure 1**

Commonly used shocks as sources of fluctuations and a stylized description of the economy overall, showing the major channels through which monetary policy shocks are transmitted to reach and affect the real sector of the economy. Specifically, monetary policy shocks can directly affect the real sector by changing firms’ wage-setting and price-setting behavior. Moreover, monetary policy can affect the real economy by changing agents’ expectations and money market interest rates, thereby having an effect on the financial sector. The transmission of the monetary policy shocks interacts with different primitive economic shocks to the financial and real sectors. Abbreviation: TFP, total factor productivity.
loop, as is emphasized by Kiyotaki & Moore (1997). Likewise, the balance sheet channel affects households, as a rise in interest rates lowers the value of their assets, reducing their demand.

The asset price channel is closely related to the balance sheet channel. Through the asset price channel, lower interest rates will cause more capital to flow into stocks and, consequently, raise the stock prices, leading to higher investment: A higher stock price generates a higher Tobin’s $q$ and makes it easier for firms to obtain outside equity financing. Note that this channel is at work even for firms that do not hold debt or assets and for which the value of their business remains unchanged. It is therefore distinct from the balance sheet channel.

Through the bank lending channel, or bank credit channel, looser monetary policy will enable banks and other monetary financial institutions to lend and provide credit more easily, which in turn will stimulate economic activity, once these lending activities come to pass. One part of the bank credit channel is essentially the balance sheet channel as applied to the operations of lending institutions. The uniqueness of the bank credit channel for monetary policy transmission is mainly a result of the special role of the financial sector in the economy relative to other sectors. Other parts of the bank lending channel concern the funds available to banks per se. Monetary policies may change the supply of loanable funds available to banks and, consequently, the total amount of credit they can extend to borrowers, including both firms and households. It has been argued that the most direct way monetary policy is able to affect the willingness and capacity of bank lending is to control the supply of bank reserves. Some scholars believe that the bank lending channel was the most important channel for monetary policy during the Financial Crisis of 2008 and the Great Recession. An important study of the bank lending channel in monetary policy was done by Bernanke (1983). Since then, there has been an extensive academic literature on the topic (e.g., Bernanke & Blinder 1988, Bernanke & Gertler 1995, Kashyap & Stein 1994).

Finally, we consider the exchange rate channel. Because monetary policy affects the quantity of domestic currency as well as the opportunity costs of holding it, it can affect exchange rates. The exchange rate channel allows exchange rate effects to change export and import decisions as well as international portfolio choices, affecting aggregate economic activity. For example, a depreciation reduces the disposable income that is left after servicing the regular payments on a foreign currency loan, since economic agents with revenues denominated in domestic currency would have to pay a greater amount following a depreciation of the currency. Exchange rates appear to move in response to many influences, however, and the relative effect of monetary policy is still ambiguous and demands more study.

2.2. A Brief History of Macroeconomic Models

According to Galí & Gertler (2007), economists and policy makers began to be skeptical about large-scale macroeconomic modeling during the 1970s for two related reasons. First, some existing models, like the Wharton econometric model and the Brookings model, failed to forecast the stagflation in the 1970s. These traditional large-scale macroeconomic models were originated by Klein (1985, 1991) and have been in use for decades. Second, leading macroeconomists leveled harsh criticisms about their underlying framework. Lucas (1976) and Sargent (1981), in particular, argued that the absence of an optimization-based approach to the development of the structural equations meant that the estimated model coefficients were likely not invariant to shifts in policy regimes, or to other types of structural changes. Similarly, Sims (1980) argued that the absence of convincing identification assumptions to sort out the vast simultaneity among macroeconomic variables meant that one could have little confidence that the parameter estimates would be stable across different policy regimes. More precisely, Sims (1980) argued that
large-scale macroeconometric models may fit the data well but that they will provide misleading answers due to noncredible identification restrictions.

Despite the criticisms by Lucas (1976) and Sims (1980), many central banks continued to use large-scale macroeconometric models and reduced-form statistical models in the 1980s and 1990s to produce forecasts of the economy that presumed no structural change, perhaps for lack of suitable alternatives. Over the past two decades, however, new quantitative and micro-founded macroeconomic frameworks for monetary policy evaluations have made large inroads. The building blocks for the development of this new framework were two independent literatures that emerged in response to the downfall of traditional macroeconometric modeling: the New Keynesian approach and the real business cycle (RBC) theory. The New Keynesian paradigm arose in the 1980s as an attempt to provide microfoundations for key Keynesian concepts such as the inefficiency of aggregate fluctuations, nominal price stickiness, and the nonneutrality of money (e.g., Mankiw & Romer 1991). The models of this literature, however, were typically static and designed mainly for qualitative, as opposed to quantitative, analysis. By contrast, RBC theory, which was developed concurrently, demonstrated how it was possible to build quantitative macroeconomic models exclusively “from the bottom up”—that is, from explicit optimizing behavior at the individual level (e.g., Prescott 1986). The RBC models usually abstracted monetary and financial factors, or assigned a negligible influence on real activity to monetary policy. New frameworks arising late in the 1990s and more prominently in the early 2000s reflected a natural synthesis of the New Keynesian and RBC approaches. A variety of labels have been used for this new framework. For example, Goodfriend & King (1997) employ the term “New Neoclassical Synthesis,” while Woodford (2003) uses “NeoWicksellian” and Clarida, Gali & Gertler (1999) use “New Keynesian.” Now, they are most often referred to as New Keynesian DSGE models, as they incorporate nominal stickiness and the resulting monetary nonneutrality into a fully specified dynamic general equilibrium framework. It is important to keep in mind, however, that a substantial portion and backbone of these models is an RBC model. Which forces dominate is then a matter of quantity, not of principle.

Today, central banks use a wide range of macroeconomic models and tools for forecasting and monetary policy analysis, including large-scale macroeconometric models, reduced-form statistical models, structural autoregressive models, and New Keynesian DSGE models. The characteristics of the various models are summarized in Table 1. Large-scale macroeconometric models constrain purely data-driven models in such a way that the long-run dynamic behavior of the variables converges to the theoretical long-run steady state. In econometric terms, the macroeconometric models developed and used by central banks are essentially large-scale restricted vector error-correction models (VECMs). This approach puts less emphasis on theory, insofar as short-run dynamics are largely data driven, and long-run relations implied by theory still have to be confirmed by empirical research. For instance, the modeler would not insist that the model has a balanced growth equilibrium, but instead would test whether the cointegrating relation implied by the model is present in the data. Examples of this type of macroeconometric model include the Bank of England’s earlier Medium-Term Macroeconometric Model (MTMM), the Bank of Canada’s Quarterly Projection Model (QPM), the Fed’s MIT–Penn–Social Science Research Council (MPS) and FRB/US models, and the ECB’s Area-Wide Model (AWM).

Although the large-scale macroeconometric model still plays an active role at major monetary authorities such as the Fed, there has been a steady shift toward models that place greater emphasis on theoretical consistency. For example, the Bank of Canada shifted its principal model from the QPM to the Terms-of-Trade Economic Model (ToTEM) in late 2005, the ECB replaced its AWM with a New Area-Wide Model (NAWM), and the Fed started to build various DSGE models such
### Table 1  Macroeconometric models, SVAR models, and New Keynesian DSGE models

<table>
<thead>
<tr>
<th></th>
<th>Macroeconometric</th>
<th>SVAR</th>
<th>DSGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example</strong></td>
<td>FRB/US, FRB/Global, AWM, MTMM, QPM</td>
<td>Linear approximation to DSGE models</td>
<td>SIGMA, EDO, CMR, ToTEM, NAWM</td>
</tr>
<tr>
<td><strong>Dynamic</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Long-run relations</strong></td>
<td>Based on steady-state equilibrium in theory</td>
<td>Based on theory and restrictions</td>
<td>Based explicitly on individual optimization in a coherent manner</td>
</tr>
<tr>
<td><strong>Short-run dynamics</strong></td>
<td>Based on ad hoc adjustment dynamics</td>
<td>Based on theory and restrictions</td>
<td>Based explicitly on individual optimization in a coherent manner</td>
</tr>
<tr>
<td><strong>Sims critique (i.e., reliable structural exogenous shocks)?</strong></td>
<td>Partly</td>
<td>Yes (ideally)</td>
<td>Yes (ideally)</td>
</tr>
<tr>
<td><strong>Lucas critique (i.e., reliable policy analysis)?</strong></td>
<td>Partly</td>
<td>Yes (ideally)</td>
<td>Yes (ideally)</td>
</tr>
<tr>
<td><strong>Policy experiment (i.e., impulse-response analysis)</strong></td>
<td>Yes (less credible)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Forecast?</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Estimated?</strong></td>
<td>Estimation and calibration</td>
<td>Yes</td>
<td>Estimation and calibration</td>
</tr>
<tr>
<td><strong>Nonlinearity?</strong></td>
<td>Maybe</td>
<td>No</td>
<td>Maybe</td>
</tr>
</tbody>
</table>

Abbreviations: AWM, Area-Wide Model; CMR, Christiano–Motto–Rostagno; DSGE, dynamic stochastic general equilibrium; EDO, estimated dynamic optimization-based model; MTMM, Medium-Term Macroeconometric Model; NAWM, New Area-Wide Model; QPM, Quarterly Projection Model; SIGMA, multicountry open economy model; SVAR, structural vector autoregression; ToTEM, Terms-of-Trade Economic Model.

As SIGMA (multicountry open economy model) and EDO (estimated dynamic optimization-based model). (Please see Supplemental Appendix B for a review of the core models used by major central banks.) This vintage of new macroeconometric models uses a calibrated theoretical model to pin down a set of steady-state attractors to describe an error-correcting relationship. Dynamics are driven by assuming that there are adjustment costs between current and long-run levels for variables on a partial equilibrium basis. Higher orders of adjustment costs introduce a role for forward-looking expectations. The full model is a mixture of structural relations implied by a partial equilibrium treatment of theory, such as the decision rule for aggregate consumption, and some reduced-form relations such as their error-correcting trade equations. Finally, SVAR models, as first introduced by Sims (1980) as an alternative to traditional large-scale macroeconometric models, enjoy a large degree of popularity at central banks.

One benefit of having multiple models is the opportunity to examine the robustness of policy strategies across models with quite different foundations. According to Tovar (2009) and Chung, Kiley & Laforte (2010), central bankers emphasize that, in their experience, model-based policy analysis is enhanced by considering multiple models, and indeed, they often learn as much when models disagree as when they agree.

In the next section, we focus on advances in the development of New Keynesian DSGE models, which now serve as core models and workhorses at several major central banks. Recent efforts on the academic side include the incorporation of financial frictions (i.e., the financial accelerator channel), financial intermediation (i.e., the bank funding channel), nontrivial fiscal policies, and...
2.3. Why DSGE Models?

DSGE models have become increasingly attractive to central banks over the past two decades. The reasons why this has happened also naturally provide answers to how we may use DSGE models. We list three of the most important reasons below.

First, DSGE models emphasize the important role of expectations in assessing alternative policy actions by relating reduced-form parameters to deeper structural parameters. This connection makes policy applications less subject to the Lucas critique (see Lucas 1976), as structural parameters are less likely to change in response to changes in policy regime.

---

7See our companion article (Dou et al. 2019) for a simple canonical New Keynesian DSGE model, of the type that has been the core component of all central bank DSGE models, which can be extended by adding an imperfect credit market and financial intermediation.
Second, impulse-response analysis allows a DSGE model to identify and decompose economic and policy structural shocks on the quantitative level. A reasonable identification of structural shocks greatly improves the reliability of policy analysis and counterfactual experiments, making the analysis less subject to the Sims critique (see Sims 1980). The nature of the DSGE model’s structure—not only in terms of its parameters but also in terms of the way exogenous shocks drive the economy according to the model—makes it possible to tell coherent stories and structure forecasts around it.

Third, the capacity of DSGE models to link model implications to time-series and cross-sectional data is particularly useful for discovering deep structural parameters. Recent advances in the construction, simulation, and estimation of DSGE models have made it possible to combine a rigorous microeconomic derivation of the behavioral equations of macroeconomic models with an empirically plausible calibration, an estimation that fits the main features of a macroeconomic time series. A series of seminal papers (e.g., Hansen & Singleton 1982, 1983; Mehra & Prescott 1985) have shown that asset pricing data are extremely useful in understanding the deep structural parameters of DSGE models. In addition, these structural parameters can be calibrated/estimated using off-model information, especially when time series are short. In terms of the accuracy of the estimation of structural parameters, DSGE models reduce the risk of overfitting by helping to identify parameters and shocks hitting the economy. The danger of overfitting returns in disguise, however, in the considerable freedom in devising such models and the particular time series used for their estimation.

An increasing number of central banks and policy institutions, including the Fed, the ECB, the Bank of England, the Bank of Canada, the Bank of New Zealand, and the International Monetary Fund, have started to use New Keynesian DSGE models as their core models. The most prominent version is the Smets & Wouters (2007) model or its earlier version (Smets & Wouters 2003), both of which are a close cousin of the Christiano, Eichenbaum & Evans (2005) model. Starting with the basic stochastic neoclassical growth model and its RBC extensions, this generation of New Keynesian DSGE models has stochastic components, real and nominal frictions such as the cost of capital adjustment, nominal wage and price rigidity, and monopolistic competition. Galí (2008) and Woodford (2003) provide an excellent introduction to the basics of New Keynesian DSGE models. A common approach used by the central banks is to start with a benchmark New Keynesian DSGE model such as the Smets & Wouters (2007) model and then to incorporate additional components, such as the following:

1. Exogenous shocks, including preference shocks, marginal efficiency shocks, global shocks, risk premium shocks, fiscal policy shocks, and so forth.
2. Frictions in the financial market, including collateral constraints, information-based frictions, moral hazard-based frictions, and limited commitment (e.g., Bernanke, Gertler & Gilchrist 1999; Brunnermeier & Pedersen 2009; Chari, Christiano & Eichenbaum 1995; Kiley & Sim 2011a,b). Other papers exploring the policy implications of both nominal rigidities and credit frictions include those on collateral-based borrowing constraints (e.g., Iacoviello 2005) and on limited access to financial markets (e.g., Galí, López-Salido & Vallés 2004, 2007).
3. Financially constrained (occasionally) intermediaries (e.g., Adrian, Moench & Shin 2010; Adrian & Shin 2010a,b; Christiano, Rostagno & Motto 2010; Gertler & Kiyotaki 2010; He & Krishnamurthy 2013).
4. Agent heterogeneity and redistribution effects of monetary policies (e.g., Algan & Ragot 2010; Auclert 2019; Eggertsson & Krugman 2012; Gornemann, Kuester & Nakajima 2012; Jermann, Schmid & Gomes 2014).

The Supplemental Appendices provide a survey of current models used by the largest central banks. In addition, in our companion article (Dou et al. 2019), we present a simple canonical model of this kind, which can be used to develop intuition and run experiments.

3. CHALLENGES AND OPPORTUNITIES FOR DSGE MODELS

Several model features and quantitative methodologies that are central to our understanding of the financial market and the macroeconomy simply are not incorporated in the standard New Keynesian DSGE models of the current generation. The Financial Crisis of 2008 and the recession that followed brought many of these missing pieces into the spotlight. It is evident that these missing pieces have a first-order impact on the economy as a whole, and have profoundly affected how governments have conducted their policies. In this section, we discuss these major missing components and methodological challenges. We hope to shed some light on the path along which researchers may advance current New Keynesian DSGE models to the next generation, making them more useful to monetary authorities. The issues of the current generation of New Keynesian DSGE models and the challenges of future improvements to these models are fundamentally and deeply interconnected. Therefore, in order to truly improve these models in one dimension, we may need to simultaneously tackle all the others to some degree.

3.1. Government Balance Sheet Irrelevance

Classic monetary macroeconomic theory, as used in modern macroeconomic models, taught in graduate school textbooks, and employed by major central banks all over the world, starts from the simple national income accounting identity, \( Y = C + I + G + X \), in which income \((Y)\) is the sum of consumption \((C)\), investment \((I)\), government spending \((G)\), and net exports \((X)\). The only role played by government in this basic framework is through government spending, the dynamics of which are specified exogenously. In other words, the effects of the government balance sheet and any intertemporal budget constraint on government are totally abstracted out of the analysis. This omission is not just some reduced-form modeling trick to simplify the analysis of monetary policy. In fact, the omission of the government balance sheet is completely justifiable in terms of both legislative practice and fundamental economic principles.

3.1.1. Legal arguments for irrelevance. From the legislative perspective, monetary policy decisions are, by law, independent of government; in other words, the fiscal anchor is independent of the monetary anchor, although the monetary anchor and the fiscal anchor inevitably interact. These monetary–fiscal interactions mainly include (a) interest rate changes, leading to changes in the “interest expense” item in the government budget, in turn leading to changes in the growth rate of government debt, which of course depends on whether taxes and expenditures react to the original changes in interest rate and, if so, by how much; (b) central banks holding earning assets (usually bonds) to back the currency they issue (which does not earn interest), giving the banks a stream of revenue (so-called seigniorage), which they generally turn over to the treasury (i.e., the government); and (c) increased inflation, reducing the real burden of the stream of future payments specified in long-term government bonds. As emphasized by Sims (2008), monetary independence could be sustained on a fair level because, up to 2007, there had been little risk on the Fed’s balance sheet. Its liabilities were mainly currency outstanding and reserve balances, and its assets were mainly short-term US Government debt.
More precisely, before 2007, there was little risk on the Fed’s balance sheet because (a) while exchange rate movements or inflation can change the value of the dollar, since assets and liabilities were all in dollars, there was no effect on net worth; (b) changes in long-term interest rates can change the market value of long bonds, but since the assets were mainly short term, this had a minor effect; and (c) the US Government was extremely unlikely to default outright on its nominal bonds, in part because, under conditions where this might be an attractive possibility, inflation to reduce the value of the debt would be easier and more efficient. Therefore, it is fair to assume that the government balance sheet played a very limited role in the Fed’s monetary policy decisions before 2007.

3.1.2. Economics arguments for irrelevance. The argument for the irrelevance of government from fundamental economic principles is more involved. The efficiency of the financial market is the key. More precisely, the financial market needs to be efficient enough that the following assumptions are satisfied:

1. Assets are valued only for their pecuniary returns. This means that assets fail to be perfect substitutes from the standpoint of investors only because of their different risk characteristics, not for any other reasons.
2. All investors can purchase arbitrary amounts of the same assets at the same market prices as the government.
3. The government conducts a Ricardian fiscal policy, indicating that the government budget constraint must be satisfied for all realizations of the price level (e.g., Woodford 1995). In the presence of multiple equilibria, a non-Ricardian spending or tax policy can trim the set of monetary policy-derived equilibria, as we discuss in Section 3.2.

Under these assumptions, the government balance sheet has no impact on the equilibrium of the economy, and hence neither does the open-market purchase of securities by the government. Thus, monetary policy models need only assume a government printing press that creates additional “money” at a greater or lesser rate, which is then put in the hands of private parties, perhaps by dropping it from helicopters. These assumptions lie at the heart of the classic monetarist view: The amount of monetary liabilities by the central bank matters for macroeconomic equilibrium, but it does not matter at all what kinds of assets might back those liabilities on the other side of the balance sheet, or how the base money gets to be in circulation.

The irrelevance or neutrality of the government balance sheet in determining market equilibrium is essentially the theoretical macroeconomic analog to the Modigliani–Miller theorem in corporate finance, as noted in the seminal paper by Wallace (1981). Wallace emphasized that this result of irrelevancy implies that both the size and the composition of the central bank or government balance sheet should be irrelevant for market equilibrium in a world with frictionless financial markets (or, more precisely, a world in which the above postulates hold). Similar to Wallace (1981), Eggertsson & Woodford (2003) derive a neutrality result in a New Keynesian model. In their framework, which assumes Ricardian fiscal policies, the portfolio of assets held by the central bank is irrelevant in determining the set of equilibrium output and price levels.

This does not, however, mean that monetary policy is irrelevant in such a world, as is sometimes thought; it simply means that monetary policy cannot be implemented through open-market

---

8The story for Europe is very different. The ECB’s assets and liabilities are denominated in different currencies because they have large noneuro reserves. Many other major economies in the world also face the same situation. In addition, it is unique to the ECB that there is no single fiscal counterpart to pressure it over seigniorage or interest expense.
operations whenever the neutrality result holds and the fiat money has zero value (e.g., Kiyotaki & Moore 2012). Control of the short-term nominal interest rate by the central bank remains possible in a frictionless environment. The central bank is still free to determine the nominal interest rate on overnight balances at the central bank. This interest rate must then be linked in equilibrium to other short-term interest rates, through arbitrage relations; thus, the central bank can determine the level of short-term nominal interest rates in general. Moreover, the central bank’s adjustment of nominal interest rates matters for the economy as a whole. Even in an endowment economy with flexible prices for all goods, the central bank’s interest rate policy can determine the evolution of the general level of prices in the economy. In a production economy with sticky prices and/or wages, it can have important real effects as well. However, even in this classic model, the effectiveness of short-term nominal interest rate policies depends heavily on the absence of arbitrage in the financial market, a condition that can be significantly violated.

The irrelevance result can be easily understood in a representative agent setting, although the result does not depend on the representative agent assumption. In representative household theory, the market price of any asset should be determined by the present value of the random returns to which it is a claim, where the present value is calculated using an asset pricing kernel, a stochastic discount factor (SDF) derived from the representative household’s marginal utility of income in different future states of the world. Insofar as a mere reshuffling of assets between the central bank and the private sector should not change the real quantity of resources available for consumption in each state of the world, the representative household’s marginal utility of income in different states of the world should not change. Therefore, the pricing kernel should not change, and neither should the market price of one unit of a given asset, assuming that the risky returns to which the asset represents a claim have not changed. More intuitively, if the central bank takes more risky securities onto its own balance sheet and allows the representative household to hold only securities that pay as much in the event of a crash as in other states, the risk does not disappear from the economy. The central bank’s earnings on its portfolio will be lower in the crash state as a result of the asset exchange, which means that earnings distributed to the treasury will be lower and, in turn, that the government will have to collect higher taxes from the private sector in that state; so the representative household’s after-tax income will be just as dependent on the risk as before. This explains why the asset pricing kernel does not change, and why asset prices are unaffected by open-market operations.

A similar result can also be derived when there are heterogeneous agents in the economy. If the central bank buys more of asset X by selling shares of asset Y, private investors should wish to purchase more of asset Y and divest themselves of asset X by exactly the amounts that undo the effects of the central bank’s trades. They optimally choose to do so to hedge the additional tax/transfer income risk that they take on as a result of the change in the central bank’s portfolio. If share \( \theta_b \) of the returns on the central bank’s portfolio are distributed to household \( b \), where \( \{ \theta_b \} \) are a set of weights that sum to one, then household \( b \) should choose a trade that cancels exactly fraction \( \theta_b \) of the central bank’s trade to afford exactly the same state-contingent consumption stream as before. Summing over all households, the private sector chooses trades that, in aggregate, precisely cancel the central bank’s trade. In fact, the representative household assumption is not essential here. As long as it is assumed that agents can fully undo the central bank’s trade, the result holds even if different households have very different attitudes toward risk, different time profiles of income, different types of nontradable income risk that they need to hedge, and so on, and regardless of how large or small the set of marketed securities may be. One can easily introduce heterogeneity of the kind that is often invoked as an explanation of time-varying risk premia without implying that any “portfolio balance” effects of central bank transactions should exist.
In fact, the portfolio balance effect is contrary to the proposition that the balance sheet size and composition are irrelevant. The portfolio balance effect of central bank transactions means that if the central bank holds less of certain assets and more of others, then the private sector is forced to hold more of the former and less of the latter as a requirement for equilibrium, and a change in the relative prices of the assets will almost always be required to induce the private parties to change the portfolios that they prefer. Therefore, portfolio balance effects imply that open market purchases of securities by the central bank must inevitably affect the market prices of those securities and, hence, other prices and quantities as well.

3.1.3. Why government balance sheets are not irrelevant. The Financial Crisis of 2008 and the Great Recession taught us that all the assumptions that guarantee the irrelevance of the government balance sheet can be violated.

First, as shown in Table 2, from October 2007 to October 2008 the size, composition, and risk characteristics of the Fed's balance sheet changed dramatically. By October 22, 2008, its assets were no longer mainly government bonds. Through the open-market purchase programs, the Fed had built up a new balance sheet with assets consisting mainly of risky loans from the private sector. These assets could potentially have suffered substantial capital losses not offset by reductions in the liabilities. On the liability side, the table shows that more than 25% of the Fed's liabilities were in the form of special deposits from the US Treasury. This made the Fed's independence fragile, and the government balance sheet began to play a potentially important role in monetary policy.

Moreover, financial market frictions apparently affected the transition dynamics of monetary policy in a nontrivial way. Indeed, there is some evidence suggesting that at least some of the Fed's special credit facilities, and similar programs of the other central banks, have affected asset prices. As a simple example, Figure 3 shows the behavior of the spreads between yields on various categories of commercial paper and the one-month overnight interest rate swap rate (essentially, a market forecast of the average federal funds rate over that horizon) over the period just before and after the introduction of the Fed's Commercial Paper Funding Facility (CPFF) at the beginning of October 2008. The reason for the introduction of the new facility was the significant disruption of the commercial paper market, indicated by the explosion of spreads in September 2008 for all four types of commercial paper shown in Figure 3. The figure also shows that spreads for three classes of paper (all except A2/P2 paper) immediately came back down with the introduction of the new facility; these three series were precisely the ones that qualified for purchases under the CPFF. In contrast, the spread for A2/P2 paper remained high for several more months, though it eventually returned to more normal levels with the general improvement of financial conditions. Spreads did not decline in the case where paper is not eligible for purchase by the new facility, suggesting that targeted asset purchases by the Fed did change the market prices of the assets.

During the Financial Crisis of 2008, conventional monetary policy measures, such as targeting a short-term nominal interest rate, had negligible effect given the zero lower bound. As a result, unconventional policy measures have become highly important. Unconventional measures that use the central bank balance sheet as an instrument include (a) changes in the supply of bank reserves beyond those required to achieve an interest rate target, (b) changes in the assets acquired by central banks (e.g., quantitative easing and credit easing), and (c) changes in the interest rate paid on reserves.

To analyze these unconventional monetary policies, we need to extend the standard New Keynesian DSGE model to allow a role for the government balance sheet in determining equilibrium, and we need to consider the connections between these alternative monetary policy measures and traditional interest rate policy. For example, Cúrdia & Woodford (2011) extended the standard New Keynesian DSGE model by allowing a transactions role for central bank liabilities.
and heterogeneous households to guarantee that the government balance sheet has a nontrivial effect on determining equilibrium. This extension allowed Cúrdia & Woodford (2011) to provide a framework to analyze unconventional monetary policy measures. In addition, Gertler & Karadi (2011) developed a quantitative monetary DSGE model with financial intermediaries that face

### Table 2  The US Federal Reserve balance sheet: assets and liabilities

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reserve bank credit</strong></td>
<td>2,154,356</td>
<td>1,803,300</td>
<td>944,345</td>
</tr>
<tr>
<td>Securities held outright</td>
<td>1,692,177</td>
<td>490,633</td>
<td>−288,947</td>
</tr>
<tr>
<td>US Treasury</td>
<td>774,552</td>
<td>476,528</td>
<td>−303,025</td>
</tr>
<tr>
<td>Federal agency and MBS</td>
<td>917,626</td>
<td>14,105</td>
<td>14,105</td>
</tr>
<tr>
<td>Repurchase agreements</td>
<td>0</td>
<td>80,000</td>
<td>42,286</td>
</tr>
<tr>
<td>Term auction credit</td>
<td>139,245</td>
<td>263,092</td>
<td>263,092</td>
</tr>
<tr>
<td>Other loans</td>
<td>107,630</td>
<td>418,580</td>
<td>418,286</td>
</tr>
<tr>
<td>Primary credit</td>
<td>22,578</td>
<td>105,754</td>
<td>105,612</td>
</tr>
<tr>
<td>Primary dealer and other broker-dealer credit</td>
<td>0</td>
<td>111,255</td>
<td>111,225</td>
</tr>
<tr>
<td>Asset-backed commercial paper money market</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Mutual Fund Liquidity Facility</td>
<td>0</td>
<td>114,219</td>
<td>114,219</td>
</tr>
<tr>
<td>Credit extended to AIG</td>
<td>42,786</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Term Asset-Backed Securities Loan Facility</td>
<td>41,818</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other credit extensions</td>
<td>0</td>
<td>87,332</td>
<td>87,332</td>
</tr>
<tr>
<td>Net portfolio holdings of Maiden Lane LLC</td>
<td>26,381</td>
<td>29,137</td>
<td>29,137</td>
</tr>
<tr>
<td>Float</td>
<td>−2,476</td>
<td>−1,048</td>
<td>−558</td>
</tr>
<tr>
<td>Central bank liquidity swaps</td>
<td>33,315</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Federal Reserve assets</td>
<td>90,476</td>
<td>522,906</td>
<td>481,050</td>
</tr>
<tr>
<td>Gold stock</td>
<td>11,041</td>
<td>11,041</td>
<td>0</td>
</tr>
<tr>
<td>Special drawing rights certificate account</td>
<td>5,200</td>
<td>2,200</td>
<td>0</td>
</tr>
<tr>
<td>Treasury currency outstanding</td>
<td>42,605</td>
<td>38,773</td>
<td>92</td>
</tr>
<tr>
<td><strong>Liability items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Currency in circulation</strong></td>
<td>913,756</td>
<td>854,517</td>
<td>41,706</td>
</tr>
<tr>
<td>Reverse repurchase agreements</td>
<td>65,737</td>
<td>98,110</td>
<td>61,384</td>
</tr>
<tr>
<td>Foreign official and international accounts</td>
<td>65,737</td>
<td>73,110</td>
<td>36,384</td>
</tr>
<tr>
<td>Dealers</td>
<td>0</td>
<td>25,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Treasury cash holdings</td>
<td>284</td>
<td>276</td>
<td>−46</td>
</tr>
<tr>
<td>Deposits with Federal Reserve banks</td>
<td>86,496</td>
<td>554,927</td>
<td>542,431</td>
</tr>
<tr>
<td>US Treasury, general account</td>
<td>43,241</td>
<td>23,166</td>
<td>18,120</td>
</tr>
<tr>
<td>US Treasury, supplementary financial account</td>
<td>29,992</td>
<td>524,771</td>
<td>524,771</td>
</tr>
<tr>
<td>Foreign official</td>
<td>2,297</td>
<td>254</td>
<td>155</td>
</tr>
<tr>
<td>Service related</td>
<td>3,237</td>
<td>6,138</td>
<td>−441</td>
</tr>
<tr>
<td>Required clearing balances</td>
<td>3,237</td>
<td>6,138</td>
<td>−441</td>
</tr>
<tr>
<td>Adjustments to compensate for float</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>7,730</td>
<td>598</td>
<td>289</td>
</tr>
<tr>
<td><strong>Other liabilities and capital</strong></td>
<td>61,537</td>
<td>46,213</td>
<td>4,273</td>
</tr>
</tbody>
</table>

---

*a* All entries are in millions of US dollars.

Abbreviations: MBS, mortgage-backed securities; NA, no data.
Figure 3

Behavior of the spreads between yields on various categories of one-month commercial paper (CP) and the one-month overnight index swap (OIS) over the period just before and after the introduction of the Fed’s Commercial Paper Funding Facility (CPFF) at the beginning of October 2008. The red line represents the spread of one-month A2/P2 CP, the gray line represents the spread of one-month AA asset-based (AB) CP, the yellow line represents the spread of one-month AA financial CP, and the black line represents the spread of one-month nonfinancial CP. Figure based on Federal Reserve economic data.

endogenously determined balance sheet constraints to evaluate the effects of the central bank using unconventional monetary policy to combat a simulated financial crisis. Unconventional tools also include forward guidance on interest rates (see Kuttner 2018 for a valuable survey). Specifically, Fed policy statements in December 2008 began to include explicit references to the likely path of the federal funds interest rate, a policy that came to be known as forward guidance.

One important channel through which the balance sheet takes effect is the intertemporal budget constraint:

\[
GL_t = \sum_{k=t}^{\infty} E_t \left[ \frac{\Lambda_{t+k}}{\Lambda_t} \left( s_k + \frac{i_k}{1+i_k} \frac{M_k}{P_k} \right) \right],
\]

where \( GL_t \) denotes the real value of net government liabilities in period \( t \); \( \Lambda_t \) is the state price density in period \( t \); \( M_k \) is the money supply; \( P_k \) is the price level; and \( s_k \) is the real primary government budget surplus, which is the difference between revenues from, on the one hand, taxes, real investments, the premium from insurance/guarantees, the assets held by the central bank and the treasury, and so forth and, on the other hand, real investments, insurance payments, and so on. Lucas (2012) reviews the theoretical and practical rationale for treating market risk as a cost to governments, presenting an interpretive review of the growing literature that applies the concepts and tools of modern finance to evaluating the costs of government policies and projects. Lucas (2012) stresses that governments typically understate their cost of capital because they identify it with their borrowing costs, rather than with a rate of return commensurate with the risk of a project.
A consequence of this understatement is that the official cost estimates for many government investment and financial activities are also understated, in some cases significantly so. However, in a few cases, risk adjustment lowers estimated costs relative to official estimates. Lucas (2012) emphasizes that when the financial market is incomplete, the choice of appropriate state price density, $\Lambda_t$, becomes critical, and in practice rather tricky. In such a complex case, different cash flows could require different SDFs. However, there is still a debate on whether the government balance sheet is constrained.

3.2. Fiscal Theory of the Price Level

Many specifications of monetary policy by themselves fail to determine a unique equilibrium in their inflation dynamics. These multiple equilibria arise when monetary policy is the sole focus of these models, the government budget constraint is ignored, and it is assumed that fiscal policy is completely accommodating to monetary policy. However, as pointed out by Leeper (1991), Sims (1994), Woodford (1994, 1995), and Cochrane (2011), an active fiscal policy will be able to trim the set of equilibria and achieve a determinate solution not only for inflation but also for the price level. We elaborate this point below using a simple model from Cochrane (2011), which we describe in detail in Supplemental Appendix A.

The equilibrium of the simple flexible price economy with a constant output $Y_t = Y$ and government spending $G_t = 0$ in each period can be summarized by the Fisher relation,

$$1 + i_t = (1 + r)\Pi_{t+1},$$

the monetary policy rule,

$$1 + i_t = (1 + r)\Phi(\Pi_t),$$

and the government's present-value budget constraint,

$$\frac{B_{t-1}}{P_t} = \sum_{j=0}^{\infty} \frac{T_{t+j}}{(1 + r)^j},$$

where $i_t$ is the nominal interest rate at time $t$, $r$ is the constant real interest rate, $P_t$ is the price level at time $t$, $\Pi_t \equiv P_t/P_{t-1}$ is inflation, $B_{t-1}$ is the gross government debt due at time $t$, and $T_t$ is the real tax at time $t$. It is assumed that $P_{-1}$ and $B_{-1}$ are given.

Combining Equations 2 and 3, we get the equilibrium path of inflation:

$$\Pi_{t+1} = \Phi(\Pi_t).$$

Clearly, in general, without an initial or terminal condition for $\Pi_t$, monetary policy by itself will not determine the inflation rate, let alone the price level. A Taylor rule implies a locally deterministic equilibrium if we ignore equilibria with explosive inflation dynamics. Figure 4 illustrates this point. If $\Pi^*$ is the desired optimal level of inflation, then a Taylor rule should have $\Phi'(\Pi^*) > 1$. However, this is an unstable equilibrium. As pointed out by Cochrane (2011), the only reason the optimal inflation, $\Pi^*$, will hold is that for any other starting $\Pi_0 \neq \Pi^*$, the monetary authority would threaten to “blow up” the economy by creating explosive inflation. Cochrane (2011) notes that ruling out such explosive equilibria has no basis in economic theory (since it is only inflation, or the nominal side, that is blowing up, not the real side, such as the asymptotic value of the debt) or in economic history (which has many recorded instances of hyperinflation). Thus, such explosive equilibria are also valid solutions in addition to $\Pi^*$. Moreover, since $i_t \geq 0$, we must have a stable equilibrium $\Pi_L$, where $\Phi'(\Pi_L) < 1$, and to which any path of inflation starting from $\Pi_0 < \Pi^*$ must converge. This adds to the multiplicity of equilibria.
All such equilibria, however, are valid only because we have ignored the government budget constraint of Equation 4 and assumed a Ricardian fiscal policy, which implies that the budget constraint must hold for all price levels and taxes must adjust to accommodate the price level obtained from monetary policy dynamics. An “active” (in the sense of Leeper 1991) or non-Ricardian (Woodford 1994) fiscal policy would force exogenous values of \{T_t\} such that only one price level given by the budget constraint of Equation 4 can hold as an equilibrium. The policy is non-Ricardian because it does not satisfy the government budget constraint for all realizations of the price level. Non-Ricardian policies trim the set of equilibria or, in this case, select one equilibrium from the many that are implied by the monetary policy. With reference to Figure 4, the non-Ricardian fiscal policy would select \( P_0 \), and inflation would subsequently follow a deterministic path to either \( \Pi_L \) or \( \Pi^* \).

To avoid explosive inflation dynamics (rather than ruling them out ad hoc), the following combination of active fiscal policy and passive monetary policy is very effective. The active fiscal policy takes the form of non-Ricardian policies, as above, while the passive monetary policy in this case means that the interest rate should react less than one-for-one to inflation, or \( \Phi'(\Pi^*) < 1 \). This would ensure stability and convergence to \( \Pi^* \) of any equilibrium path, the initial \( P_0 \) of which would be determined by the active fiscal policy (Figure 5).

The multiplicity of equilibria is a common feature in New Keynesian models. When monetary policy is unable to determine the equilibrium alone, the key to determining a unique equilibrium is the specification of additional policies to the monetary policy. Optimal policy prescriptions in zero-lower-bound situations (e.g., Werning 2011) must rely heavily on the Taylor rule to select the right equilibrium path. The Taylor rule, however, generates uniqueness only locally, not globally, as shown in Figure 4, and only after ruling out ad hoc explosive inflation dynamics.

This begs the question of whether a non-Ricardian policy for long-term interest rates coupled with a passive monetary policy can similarly be used to select equilibria. Quantitative easing—if

---

**Figure 4**

Inflation dynamics with a Taylor rule. The x-axis is the inflation rate in period \( t \), and the y-axis is the inflation rate in period \( t + 1 \). Figure adapted from Cochrane (2011).
we interpret it as an active policy for long-term interest rates—can help select the appropriate nondeflationary equilibrium. The crux of the ability to set long-term interest rates independently of the short-term rates can be seen only in a nonlinearized model with uncertainty, with a policy for short-term rates that does not depend on the realization of any of the endogenous variables. Changing the long-term interest rate independently of the short-term rate implies reweighting the future short-term rates by a different SDF in an economy with variable output. We believe this is an important question to address that has significant implications for policy making.

3.3. Heterogeneity, Reallocation, and Redistribution Effects

In the core monetary models employed by major central banks, the existence of a representative agent is assumed. It has been widely recognized that important features of macroeconomic data are difficult or impossible to explain within the representative agent framework. These include the cyclical behavior of the factor shares of national income; the large risk premium (e.g., Basak & Cuoco 1998, Guvenen 2009); cross-sectional stock returns (e.g., Fama & French 1992); the distribution of income, wealth, and leverage of households; and the distribution of leverage, asset, and cash holdings in firms. The data also show strong cyclical patterns of capital, labor, and credit reallocation among firms, along with cyclical behavior in the bankruptcy rates, entries and exits, and mergers and acquisitions among firms (e.g., Eisfeldt & Rampini 2006, 2008). All of these behaviors should have significant equilibrium effects on aggregate quantities. An important feature of the data is that households do not mainly save by themselves to finance their consumption; instead, one side of households finances the other side’s consumption (e.g., Guerrieri & Lorenzoni 2012).

Theoretically, the representative agent can be justified by the assumption of complete markets in an economy without externality. Complete market conditions imply perfect insurance for agents in the economy, or that insurance in the economy is costless. However, in reality, the profits
of insurance companies in the USA and costly transactions and portfolio constraints demonstrate the violation, or at least the poor approximation, of the complete market assumption. Also, under assumptions like a time-separable utility function and the stationarity of the economy, dynamically complete market models predict the capital structure of firms and the asset portfolio of households to be constant, even when the investment opportunity varies over time and there are heterogeneous agents (e.g., Judd, Kubler & Schmedders 2003). This prediction is largely contradicted by the real data. Moreover, Sims (2006) argues that there is no overall aggregate capital and no aggregate consumption good, and that the real economy has a rich array of financial markets, which models have so far not included in a widespread or successful manner. Aggregate identities are problematic when we realize that different agents in the economy deal with uncertainty and risk differently.

There are three main reasons for incorporating rich heterogeneity into macroeconomic models. First, it is important not to ignore the significant equilibrium effects of the distribution of wealth, income, leverage, cash holdings, and so forth on the aggregate quantities and transitional dynamics of various monetary policies (e.g., Guerrieri & Lorenzoni 2012). Moreover, from first principles, the cross-sectional distribution serves as an infinite-dimensional important endogenous state variable, so it has a strong impact on equilibrium. Second, heterogeneity makes it possible to analyze the equilibrium effect of extensive margins on the aggregate quantities and transitional dynamics of monetary policies. Third, the heterogeneity of agents provides a framework to assess the welfare properties of different monetary policy measures, fiscal policies, and other government policies, including unemployment insurance and social security programs.

The key feature of models with heterogeneous agents that makes them different from representative agent models is that the set of possible trades available for agents is restricted. The trading restrictions are usually modeled as an incomplete set of Arrow–Debreu securities, portfolio constraints, or trading frictions, which prevents various aggregation results from holding (e.g., Deaton 1992). For example, Constantinides & Duffie (1996) show that the heterogeneous agent model with incomplete markets can be represented as several homogeneous agent models as long as the shocks in the model are such that agents do not gain anything by trading, even in the presence of those markets. Therefore, computing the equilibrium requires keeping track of the distribution of agents. In each time period, the state of the economy is characterized by exogenous state variables driven by exogenous shocks, and by endogenous state variables whose laws of motion are endogenously determined in the economy. The endogenous state variables usually include the distribution of agents. The difference between these distributions and distributions from ordinary endogenous state variables such as the capital stock is that they are usually infinite-dimensional mathematical objects. The equilibrium prices and quantities are functions of the (potentially infinite-dimensional) endogenous and exogenous state variables. The laws of motion for endogenous state variables are usually called the transition map, and because they are determined in equilibrium, it is intrinsically difficult to solve the transition map for an infinite-dimensional vector of state variables. The aggregate level of prices and quantities is not enough to characterize the state of the economy or predict future endogenous state variables, which is the key feature distinguishing the economy of heterogeneous agents and incomplete markets from the representative agent economy with complete markets. As Hall (1978, p. 974, corollary 1) states, “[n]o information available in period t apart from the level of consumption, \(c_t\), helps predict future consumption, \(c_{t+1}\), in the sense of affecting the expected value of marginal utility. In particular, income or wealth in period t or earlier [is] irrelevant, once \(c_t\) is known.”

Solving the heterogeneous agent model entails solving the policy functions and the transition map simultaneously. Mathematically, this amounts to solving a fixed-point problem for an infinite-dimensional object. The standard numerical methods include discretization of the state
space, parameterization of distributions, backward and recursive methods, and so on. To develop a sense of the extra computational complexity caused by an endogenous transition map, note that solving the agent’s problem for a given transition map of endogenous state variables is not enough. The correct transition map has to be found at the same time. This requires a double-layer iteration algorithm. To circumvent this requirement, or to avoid iterations on the transition map, one needs to prevent the distribution of agents from affecting relative prices, which dramatically simplifies the computations. For example, Aiyagari (1994) focuses on the steady state of an economy without aggregate fluctuations. In similar fashion, Imrohoroglu (1989) utilizes a storage technology that pins down the rate of return of savings exogenously, while Diaz-Gimenez (1990) assumes that the government commits itself to a specific inflation rate policy that does not depend on the asset distribution.

3.3.1. Heterogeneity and New Keynesian models. Heterogeneous agent models have lately been incorporated into the New Keynesian DSGE framework to study the effects of monetary policy. Algan & Ragot (2010) show the importance of a new precautionary savings motive in an incomplete market model in which the traditional redistributive effects of inflation are also introduced. They also show that the long-run neutrality of inflation on capital accumulation obtained in complete market models does not hold under household binding credit constraints. They demonstrate that there is a quantitative rationale for the observed hump-shaped relationship between inflation and capital accumulation.

Borrowing-constrained households are not able to adjust their money holdings differently compared with unconstrained households since they cannot rebalance their financial portfolio when fluctuations in inflation become large. Inflation, therefore, increases capital accumulation as a result of the precautionary saving motive under heterogeneity.

It is necessary to understand heterogeneity to better study the redistribution effects of monetary policy. Gornemann, Kuester & Nakajima (2012) show that heterogeneous workers vary in their employment status due to search and matching frictions in the labor market, their potential labor income, and their amount of savings. Their New Keynesian model quantitatively assesses who stands to gain or lose from unanticipated monetary accommodation, and who benefits the most from systematic monetary stabilization policy. They find substantial redistribution effects from monetary policy shocks. A contractionary monetary policy has opposing effects on the wealthiest 5% versus the rest of the population. The top 5% enjoy increases in income and welfare, while the remaining 95% suffer under a contractionary monetary policy shock. Consequently, the negative effect of a contractionary monetary policy shock to social welfare is larger if heterogeneity is taken into account. In an influential paper, Eggertsson & Krugman (2012) provide a theoretical discussion of the importance of the redistribution effect of monetary policy between creditors and debtors in understanding the economic difficulties during the period 2007–2009. As for the redistribution effect of monetary policies, firm heterogeneity is also important when it comes to inflation dynamics, investments, and risk premia (e.g., Jermann, Schmid & Gomes 2014).

3.3.2. Heterogeneity is not new in macroeconomics. The effects of heterogeneity have long been studied in macroeconomics, leading to its serious adoption in New Keynesian DSGE models. Imrohoroglu (1989) may have authored the first published paper to compute the equilibrium of a model with heterogeneous agents, and to calibrate it to match key US observations. Imrohoroglu (1989) considers different institutional market arrangements under three different environments and evaluates the welfare difference across institutional market arrangements. Similar welfare levels indicate that the existence of liquidity constraints in an economy is trivial for welfare considerations. Diaz-Gimenez (1990) explores the business cycle implications of
alternative insurance technologies using a methodology similar to Imrohoroglu’s, which could be easily adjusted to study the welfare effects of monetary and fiscal policy. Diaz-Gimenez (1990) compares perfect insurance and monetary arrangements with pervasive liquidity constraints, finding that the welfare costs of monetary arrangements were 1.25% of output in zero-inflation economies. Hansen & Imrohoroglu (1992) find that the optimal level of unemployment insurance is very low, even when there is a very small amount of moral hazard.

Huggett (1993) explains the puzzle of very low risk-free interest rates in the postwar period in the USA by assessing the importance of the role played by the lack of insurance. Huggett’s model does not have aggregate uncertainty and assumes an economy in which agents, subjected to idiosyncratic labor market shocks of the same type studied by Imrohoroglu (1989), can lend and borrow up to certain limits at a rate that is endogenously determined by nontrivial market-clearing conditions, which are necessary to solve for the equilibrium of this economy.

Aiyagari (1994) describes two features: first, an endowment economy that has no possibilities to save as a whole, and second, the level of aggregate savings affecting the society’s ability to produce goods. Aiyagari (1994) incorporates these features by using the standard neoclassical growth model with production. In order to measure the size of the role of precautionary savings, especially those motivated by self-insurance against idiosyncratic risk, Aiyagari (1994) has to deviate from the endowment economy setting from Huggett (1993). Aiyagari (1994) finds that with moderate and empirically plausible parameter values, uninsured idiosyncratic risk accounts for a 3% increase in the aggregate savings rate.

Krusell & Smith (1998) propose an important method for solving models with heterogeneity and aggregate uncertainty. When there are aggregate shocks in the model, the entire wealth distribution is an endogenous state variable, but its distribution can be approximated by its first few moments. The authors find that this approximate aggregation is reasonable, and to forecast future prices and quantities, it is enough to use the mean wealth instead of the entire cross-sectional wealth distribution. The distribution of wealth is unimportant to aggregate quantities such as aggregate consumption when most agents have the same marginal propensity to consume after aggregate shocks. Most agents achieve good self-insurance in the model, which is equivalent to saying that the consumption policy functions are roughly linear. Aggregate capital by design is three times larger than output; therefore, most agents are rich enough to almost completely smooth out shocks. Only very poor agents, who account for a small fraction of aggregate consumption, do not have self-insurance. Krusell & Smith (1998) conduct an experiment to compare the model under complete and incomplete market conditions, and find that heterogeneity has little effect on the model’s business cycle properties.

3.3.3. Liquidity and heterogeneous firms. It is well known that efficient trade and the reallocation of resources among different agents and sectors have a substantial impact on the macroeconomic performance and transitional dynamics of monetary policy (e.g., Walsh 2012). However, the data show that resource mobility is far from frictionless, and the intensity of resource reallocation has strong cyclical patterns (e.g., Eisfeldt & Rampini 2006). The imperfect nature of resource mobility plays a surprisingly small role in most policy models in use at major central banks. In those core New Keynesian DSGE models, for example, it is costly for firms to adjust their selling prices, but those same firms can hire and fire workers without cost, and both workers and capital can frictionlessly shift from one firm to another.

Theoretically, ignoring the potential costs associated with shifting real economic resources is consistent with a standard economy with one sector of homogeneous firms and representative households. However, real-world economies consist of multiple sectors and heterogeneous agents, and the data show that different sectors of the economy and different firms and households behave...
very differently over the course of the business cycle. For example, durable goods-producing sectors are more cyclically sensitive than service sectors. Economic fluctuations may be associated with shifts in relative prices across sectors, or with persistent shifts in relative demand that may require labor and capital to shift from contracting to expanding sectors of the economy and from low-productivity firms to high-productivity firms. These shifts require resources to transfer, yet differences in labor skills or in the type of capital employed in different occupations or sectors may make sectoral reallocations costly. The costs that arise because resources are not fully mobile may have consequences for policies on aggregate demand. Monetary policy shocks will definitely alter the transitional dynamics of the demand shock. For example, Walsh (2012) concludes that resource mobility matters for both the transitional dynamics of monetary policy shocks and the goals of monetary policy. Resource mobility affects the transmission mechanism that links monetary policy instruments to inflation and the real economy, thereby affecting the trade-offs faced by the policy authority and the way policy makers weigh their objectives.

One important type of resource reallocation is capital reallocation. Eisfeldt & Rampini (2006) define the ease of capital reallocation between firms as capital liquidity and show that the amount of capital reallocation between US firms is procyclical. In contrast, the benefits to capital reallocation appear to be countercyclical. The benefits of capital reallocation are approximated by the dispersion among the productivity of firms. This is intuitive because smart capital should flow out of low-productivity firms into high-productivity firms. Eisfeldt & Rampini (2006) document that capital mobility is far from frictionless and particularly difficult in bad economic times. In order to quantify the cost of capital reallocation, they calibrate a simple model economy in which capital reallocation is subject to a standard adjustment cost function and impute the cost of reallocation. They find that reallocation costs need to be substantially countercyclical to be consistent with the observed joint cyclical properties of reallocation and productivity dispersion. Eisfeldt & Rampini (2008) provide a possible microfounded explanation for this endogenous inefficient capital reallocation. They argue that when managers have private information about the productivity of assets under their control and receive private benefits, substantial bonuses are required to induce less productive managers to declare that capital should be reallocated. Capital is less productively deployed in downturns because agency costs make reallocation more costly.

Another important type of resource reallocation is labor reallocation. Research by Davis, Haltiwanger & Schuh (1998) has been central to the surge of interest in this area. Their empirical analysis is based on data for manufacturing plants covering the period from the early 1970s to the mid-1980s. After defining employment increases at new and growing plants as job creation, and decreases at dying and shrinking plants as job destruction, they point out a number of empirical regularities. One striking feature is that the data are marked by a high rate of job creation and destruction. On average, close to 1 in 10 manufacturing jobs disappeared in a given year, while the rate of new job creation was slightly lower. These changes are quite persistent: A year later, nearly 7 out of 10 newly created jobs were still in existence, and approximately 8 in 10 lost jobs were still lost. In addition, job creation and destruction tended to be concentrated at plants that experienced large changes in employment (those associated with plant shutdowns and startups, for instance). Another finding is that job destruction varied more noticeably over the cycle than job creation. The data show that job destruction tended to increase sharply during a recession and then fall back, while job creation did not move as much.

Some questions have been raised about these results. For instance, some economists have cautioned against relying on data for a single sector of the economy, especially manufacturing, where employment has been shrinking so noticeably. Furthermore, the data cover a relatively limited span (the 1970s and the 1980s), and it is possible that the recessions of this period fundamentally differ from previous (or subsequent) recessions in terms of restructuring and
reallocation. Though the issue is not yet settled, some of the findings by Davis, Haltiwanger & Schuh (1998) have been replicated elsewhere. For instance, Blanchard & Diamond (1990) relied mainly on data from the Current Population Survey, which is not restricted to manufacturing alone, and discovered the same finding about the relative volatility of job creation and destruction. For example, they find that “booms are times of low job destruction rather than high job creation” (Blanchard & Diamond 1990, p. 87). Similar patterns have been discovered in data for foreign countries as well.

Kuehn, Petrosky-Nadeau & Zhang (2012) argue that frictions in the labor market are important for understanding the equity premium in the financial market. The authors embed the Diamond–Mortensen–Pissarides search framework into a DSGE model with recursive preferences. The model produces realistic equity premium and stock market volatility, as well as a low and stable interest rate. In particular, the authors show that in their model the job flows and matching friction can help generate disasters in employment, output, and consumption along the lines of Rietz (1988) and Barro (2009). Moreover, when incorporated into otherwise standard RBC models, the search framework significantly improves their empirical performance. More importantly, it allows one to analyze the cyclical behavior of unemployment, job vacancies, and job flows, important phenomena that general equilibrium models based on Walrasian labor markets are not designed to address. For example, Merz (1995) tries to explain some cyclical behavior in the US labor market by introducing a two-sided search into the labor market as an economic mechanism propagating technological shocks into a standard business cycle model. Andolfatto (1996) shows that the labor market search is a quantitatively important propagation mechanism in generating business cycles. den Haan, Ramey & Watson (2000) emphasize the economic importance of the interaction between the capital adjustment cost and the labor destruction rate in propagating technology shocks. Gertler & Trigari (2009) extend period-by-period Nash bargaining to staggered multiperiod wage contracts and show that it can account for the volatile behavior of labor market activities. Hall (2005) generates endogenous wage stickiness under a matching framework and shows that sticky wages, in turn, make labor market activities realistically sensitive to aggregate shocks.

Given the significant equilibrium effects of job market reallocation, it is reasonable for us to speculate that job market mobility should have an important impact on the transitional dynamics of monetary policy shocks. In fact, there is an extensive literature that focuses on the positive implications of labor market friction in New Keynesian models, that is, how search and matching frictions affect the empirical performance of the New Keynesian model and the transitional dynamics of monetary policy (e.g., Cheron & Langot 2000, Christoffel & Linzert 2005, Krause & Lubik 2007, Moyen & Sahuc 2005, Trigari 2006, Walsh 2005). Thomas (2008) analyzes the optimal monetary policy under the New Keynesian framework with search and matching frictions. Monetary policy shocks should affect job market flows in a nontrivial way.

Finally, the most important resource reallocation is the reallocation of credit or funding among firms or agents. The reallocation of funding is key, partly because it can possibly explain capital reallocation and labor reallocation, as is discussed by Eiffeld & Rampini (2006). According to Bernanke, Gertler & Gilchrist (1999) and Kiyotaki & Moore (1997), the representative firm or the productive agent is impatient enough that the firm or agent does not save very much and does not escape its financial constraints. Consequently, the models have two salient features. First, the firm or the agent saves by itself and uses the savings to invest later. Second, the economy as a whole is financially constrained in the steady state. However, both implications are inconsistent with the data. In contrast, the data suggest that only a fraction of firms are occasionally bound by financial constraints, and that firms also finance one another’s investment. Fund reallocation among firms is one of the key functions of the financial sector.
3.4. Risk Premium Dynamics

Does the risk premium matter for macroeconomic dynamics and the transitional dynamics of monetary policy? The key features for risk premium dynamics include high levels of volatility, nonlinearity, and countercyclicality. However, as explained by Jermann (1998), Lettau & Uhlig (2000), and Kaltenbrunner & Lochstoer (2010), it is often difficult to generate endogenously a large and time-varying market price of risk in a production economy.
Rouwenhorst (1995) shows that the standard RBC model fails to explain the equity premium because of consumption smoothing. Using models with internal habit preferences, Jermann (1998) and Boldrin, Christiano & Fisher (2001) use capital adjustment costs and cross-sector immobility, respectively, to restrict consumption smoothing to explain the equity premium. However, both models struggle with excessively high interest rate volatilities. Using recursive preferences to curb interest rate volatility, Tallarini (2000) and Kaltenbrunner & Lochstoer (2010) show that baseline production economies without labor market frictions can explain the Sharpe ratio, but still fail to match the equity premium and the stock market volatility.

Uhlig (2007) shows that wage rigidity helps explain the Sharpe ratio and the interest rate volatility in an external habit model, but the equity premium and the stock market volatility are close to zero in this model. Gourio (2011) shows that operating leverage derived from labor contracting helps explain the cross section of expected returns, but does not study aggregate asset prices. Favilukis & Lin (2012) quantify the role of infrequent wage renegotiations in an equilibrium asset pricing model with long-run productivity risk and labor adjustment costs. They argue that, in standard models, highly procyclical and volatile wages are a hedge against adverse shocks of productivity for the shareholder. The residual—profit or dividends—becomes unrealistically smooth, as do returns. Smoother wages act like operating leverage, making profits more risky. Bad times and unproductive firms are especially risky because committed wage payments are high relative to output. Instead of specifying the wage rule exogenously, Kuehn, Petrosky-Nadeau & Zhang (2012) differ from the abovementioned authors by using the search framework to derive equilibrium wages. Because dividends equal output minus wages minus total vacancy costs (in a manner analogous to investment), providing a microfoundation for equilibrium wages makes the dividends truly endogenous in a production economy.


Gourio (2012) shows that an increase in disaster risk leads to a decline in employment, output, investment, stock prices, and interest rates, as well as an increase in the expected return on risky assets. The model matches the data well on quantities, asset prices, and particularly the relationship between quantities and prices, suggesting that variation in aggregate risk plays a significant role in some business cycles. More precisely, the mechanism is that an increase in the disaster probability affects the economy by lowering expectations and increasing risk. Because investors are risk averse, this higher risk leads to higher risk premia, with significant implications for both business cycles and asset prices: Stock prices fall, employment and output contract, and investment, especially, declines. Demand for precautionary savings increases, leading the yield on less risky assets to fall, while expected excess returns on risky securities increase. These dynamics occur in the absence of any change in total factor productivity (TFP).

Risk premia are important in understanding many macroeconomic questions, for instance, why investment is often low despite low riskless interest rates. Here, the relevant user cost of capital may well be high if the riskless interest rate is low precisely because of high disaster risk. This will directly affect the transitional dynamics of monetary policies.

Introducing time-varying risk requires solving a model using nonlinear methods, that is, going beyond the first-order approximation and considering higher-order terms in the Taylor expansion. Researchers disagree on the importance of these higher-order terms, and a fairly common view is that they are irrelevant for macroeconomic quantities. In his presidential address to the American Economic Association, Lucas (2003, p. 7) summarizes this perspective:

---

Access provided by 108.7.222.236 on 12/07/20. For personal use only.
Tallarini uses preferences of the Epstein–Zin type, with an intertemporal substitution elasticity of one, to construct an RBC model of the U.S. economy. He finds an astonishing separation of quantity and asset price determination: The behavior of aggregate quantities depends hardly at all on attitudes toward risk, so the coefficient of risk aversion is left free to account for the equity premium perfectly.

However, Gourio (2012) shows that when the risk is large and varies over time, risk aversion affects macroeconomic dynamics in a significant way. In a similar spirit, but using a two-country open economy setting, Dou & Verdelhan (2014) show that the time-varying risks generate rich joint volatile dynamics of international asset prices and capital flows.

The following are some specific examples of the potential importance of the time-varying risk premium on macroeconomic dynamics and the transitional dynamics of monetary policy. Gilchrist & Zakrajsek (2012) show that the default premium, rather than the default probability, is the informative variable about macroeconomic conditions. Gilchrist, Sim & Zakrajsek (2010) show that an uncertainty shock can boost the default premium strongly, without increasing the default probability of firms significantly. The extremely high default risk premium prevents firms from investing optimally, even when the risk-free rate is low. The term premium is crucial to accurately characterize the aggregate demand relationship (the IS curve). According to Gali & Gertler (2007), the aggregate demand depends on the gap between the long-term interest rate and its natural correspondence in a model economy with flexible prices. The relationship between the long-term and short-term interest rates is captured by the term premium, which depends on the risk premium of investors. According to Gali & Gertler (2007), the IS curve is also characterized by the relationship between aggregate demand and marginal $q$. The dependence between marginal $q$ and the short-term interest rate is also largely captured by financial friction and the risk premium. However, to the best of our knowledge, generating a realistic term premium is still a challenging task in a model production economy. Reasonable risk premia, including currency and sovereign risk premia, are central to understanding international financial linkages and capital flow dynamics, which, in turn, have a nontrivial impact on the implications of monetary policy.

Finally, as emphasized in Section 2.3, the availability of rich financial data makes DSGE models particularly useful in learning deep structural parameters. However, the absence of reasonable risk premium dynamics in DSGE models wastes the information embedded in asset prices.

### 3.5. Uncertainty

The uncertainty shock has an adverse effect on macroeconomic quantities and can even drive business cycles. For example, the Federal Open Market Committee minutes repeatedly emphasize uncertainty as a key factor driving the 2001 and 2007–2009 recessions, while Stock & Watson (2012, p. 26) conclude that “[t]he main contributions to the decline in output and employment during the [2007–2009] recession are estimated to come from financial and uncertainty shocks.” Also, in an early seminal paper, Romer (1990) emphasizes the macroeconomic importance of uncertainty shocks.

In addition, economists have empirically determined that the two most important shocks that drive aggregate fluctuations are “financial disruption” and “heightened uncertainty” (e.g., Christiano, Rostagno & Motto 2010, Del Negro & Schorfheide 2012). However, it is important to have a better understanding of the sources of financial and uncertainty shocks in macroeconomic models and their endogenous interactions.

In fact, there has been a fast-growing literature studying the aggregate effects of such uncertainty shocks (e.g., Arellano, Bai & Kehoe 2011; Bachmann & Bayer 2014; Bloom 2009; Bloom et al. 2013; Bundick & Basu 2014; Christiano, Rostagno & Motto 2010, 2014; Gilchrist, Sim & Zakrajsek 2010; Herskovic et al. 2014; Pástor & Veronesi 2006, 2009). Note that the use of the
term “uncertainty” here is different from “Knightian uncertainty,” which emphasizes situations where agents cannot know all the information they need to set accurate odds (e.g., Hansen & Sargent 2008, Knight 1921). In addition, the use of “uncertainty” here is also different from aggregate volatility, which has also been extensively studied in the literature (e.g., Ai & Kiku 2016; Bansal & Yaron 2004; Campbell, Giglio & Polk 2013; Campbell et al. 2015; Drechsler & Yaron 2011; Fernández-Villaverde et al. 2011; Gourio, Siemer & Verdelhan 2015; Nakamura, Sergeyev & Steinsson 2014; Segal, Shaliastovich & Yaron 2013; Shaliastovich 2015).

3.5.1. Impact of uncertainty shocks. Since the Financial Crisis of 2008 and the Great Recession, policy authorities and academic researchers have engaged in a vigorous debate on the impact of uncertainty shocks on the joint dynamics of macroeconomic quantities and asset prices. Policy authorities, including the Fed and the ECB, have claimed that uncertainty has had an adverse effect on their economy, and have built uncertainty shocks into their core DSGE models as a main driver of aggregate fluctuations (Christiano, Rostagno & Motto 2010, 2014). Moreover, there is an extensive academic literature showing the adverse effect of uncertainty (e.g., Basu & Bundick 2011; Bloom 2009; Bloom et al. 2013; Gilchrist, Sim & Zakrajsek 2010).

However, there are two major concerns with this narrative. First, the causal relationship between fluctuations in uncertainty and fluctuations in the economy is far from clear to policymakers and researchers. Although the correlation between fluctuations in uncertainty and the economy is evident, it is still undetermined whether the heightened uncertainty partially caused the Great Recession, and whether it should be blamed for prolonging the recovery process out of the Great Recession. This is due to both the lack of crystal-clear empirical evidence and the lack of comprehensive theoretical studies on the equilibrium feedback effect between fluctuations in uncertainty and the economy, as suggested by Bloom (2013) in a review on uncertainty. Second, it has been argued that uncertainty could have a positive effect on investment and the stock market. Pástor & Veronesi (2006) use a simple calibrated stock valuation model with uncertainty to show that the fundamental value of a firm increases with uncertainty about its average future profitability, and this uncertainty was extremely high in the late 1990s. Bar-Ilan & Strange (1996) show that in a high-uncertainty environment, the benefits from investment, including the growth opportunity caused by investment lags and abandoned project options, can dominate the cost of investment, the loss of the real option value of waiting. As a result, high uncertainty can sometimes promote investment.

In fact, there is a rich, if contradictory, literature on the relationship between uncertainty and macroeconomic quantities including consumption and investment. Different theories emphasize different channels, some showing a positive relationship and others showing a negative relationship. As a whole, the impact of uncertainty is still ambiguous. The basic channels under consideration include the real option channel (i.e., the option to wait), the risk premium channel, the precautionary savings channel, the growth opportunity channel, the Oi–Hartman–Abel–Caballero channel, and the learning-by-doing channel.

The first channel under consideration, the real option channel, appears to be the most direct means by which uncertainty can potentially affect a firm’s investment and hiring decisions. The idea is that the sizable adjustment cost in investment and hiring (e.g., Cooper & Haltiwanger...
2006, Ramey & Shapiro 2001) and its irreversibility (e.g., Kogan 2001, Pindyck 1991) together make the investment decision effectively a decision on exercising call options. This real option can be viewed as an option to wait, and the opportunity cost of delay is the forgone income from the project, which is unaffected by uncertainty. This asymmetric effect of uncertainty on the benefits and costs of waiting captures the essence of the real option effect, which Bernanke (1983) refers to as the “bad news principle.” However, the real option effect can be alleviated or even overturned when environmental variables shift. For example, when projects have a liquid reallocation market (i.e., reversible), the real option effect is negligible. Another, more relevant, example is when firms are financially constrained. As demonstrated by Bolton, Wang & Yang (2013), for financially constrained firms, the uncertainty shock can have both a positive and negative effect on a firm’s investment and financing decisions.

The idea behind the second channel, the risk premium channel, is that uncertainty reduces aggregate investment, hiring, and growth through a sharp increase in the risk premium. The risk premium channel plays a key role in linking asset pricing to the interaction between uncertainty and investment/hiring, an idea that has been missing in the uncertainty literature, although with a few exceptions (e.g., Arellano, Bai & Kehoe 2011; Christiano, Motto & Rostagno 2014; Gilchrist, Sim & Zakrajsek 2010). The key idea is that in an economy with corporate debt and costly default, higher uncertainty raises the default probability for those firms that are already near default boundaries, so the cost of debt financing increases. This, in turn, reduces investment, increases the default probabilities for firms originally further from the default boundaries, and, accordingly, diminishes hiring, which then leads to lower consumption of households. This adverse feedback loop causes a ripple effect, dragging the whole economy into recession while creating sky-high credit spreads. It is clear that if financial intermediaries are strong, and very few firms are close to their financial binding constraints, the risk premium effect on an economic downturn will largely be dampened. This is a nontrivial point in generating rich and realistic endogenous uncertainty dynamics.

The third channel, the precautionary saving channel, focuses on households. It is evident that higher uncertainty depresses household consumption expenditures (e.g., Bansal & Yaron 2004). In a full-closed economy, the motivation to increase precautionary savings will also reduce contemporaneous consumption and, at the same time, increase investment. However, the investment will also drop when uncertainty is high, assuming price rigidity (e.g., Basu & Bundick 2011, Leduc & Liu 2012).

Growth opportunities, the fourth channel we consider, are the major force generating a positive association between uncertainty and investment. This idea is usually implemented in two ways in the literature. Following Bar-Ilan & Strange (1996), the first method assumes that there is an investment lag with a time to build, \( b > 0 \), and an abandonment option available for each project. The abandonment option means the loss is bounded below in bad states, while the time-to-build feature forces the firm to invest earlier in order to be able to capture opportunities in the near future. The two components together cause the rational firm to invest sooner in a high-uncertainty environment. The second method is to model two capital goods: traditional capital, called “trees,” and investment options, called “seeds” (e.g., Jovanovic 2009). In a high-uncertainty environment, the investment in seeds experiences a gradual boom.

The fifth channel, the Oi–Hartman–Abel–Caballero channel, is based on research by Oi (1961), Hartman (1972), Abel (1983), and Caballero (1991). The key idea of these models is that the adjustment cost of capital makes investment less flexible than labor adjustment. This concept, combined with a constant-return-to-scale technology, makes the marginal product of capital a convex function of output price. It follows from Jensen’s inequality that uncertainty in output price leads to a high marginal product of capital and, hence, to a high intensity of investment.
Finally, the learning-by-doing channel assumes that investors or firms have imperfect information about the underlying state of the economy, and that the only way to receive extra signals about its true state is by a sequence of investments. It naturally follows that in a high-uncertainty environment, firms conduct earlier and more intensive investment to learn the true state (e.g., Pavlova 2002, Pindyck 1993, Roberts & Weitzman 1981).

An important and still unanswered question is: Which channel dominates under which economic conditions? It is possible that the sign and magnitude of the impact of uncertainty shocks on investment and asset prices depend on the soundness of the financial system and the prevailing external financing costs. When financial intermediaries are strong and the risk premium is low, negative effect channels such as the real option channel and the risk premium channel will have limited impact because investment options are deep out of the money, and it is hard to trigger a crash in the financial market. In contrast, positive effect channels are given full play in this environment. Therefore, higher uncertainty should lead to earlier and more intensive investment and create a stock market boom. When the financial intermediaries are fragile, however, the real option channel with liquidity hoarding\textsuperscript{11} and the risk premium channel will dominate, while positive effect channels will play a very limited role.

3.5.2. Measuring uncertainty. It is impossible to measure uncertainty directly given that it is unobservable, conceptual, and ex ante. After all, it lives in people’s minds and has no direct material instantiation. Therefore, a range of proxies have been employed to study the impact of uncertainty shocks. Aggregate market volatility and aggregate TFP volatility are among the most popular proxies for uncertainty shock in the existing literature (e.g., Bansal et al. 2014, Bloom 2009, Campbell et al. 2012). These aggregate volatility proxies are usually referred to as “macro uncertainty.” The uncertainty shock is also approximated by an increase in the cross-sectional dispersion among agents (e.g., Bloom 2009, Gilchrist, Sim & Zakrajsek 2010). These dispersion-based measures are referred to as “micro uncertainty.” There are also measures based on survey data, including forecaster disagreement and news mentions of uncertainty. Empirically, they are all believed to be reasonably good proxies because they comove over time. This comovement itself is a nontrivial puzzle to solve, and its solution should shed light on the development of a better proxy for uncertainty and, therefore, a better equilibrium impact of uncertainty shocks.

3.6. The Financial Sector and Systemic Risk

Systemic risk is believed to be a key driver of the Financial Crisis of 2008 and the Great Recession. Rooted in the financial sector, and magnified via contagion, systemic risk has a strong adverse effect on the whole economy. However, the key players in systemic risk—financial intermediaries—are missing from the New Keynesian DSGE models used by central banks. Incorporating the financial sector into these models is necessary for studying endogenous systemic risk in the economy and analyzing conventional and unconventional monetary policies and macroprudential policy. After all, the monetary authority’s primary task is to maintain a healthy financial system in normal times and restore a distressed financial system in times of crisis.

Two critical effects must be considered when incorporating financial intermediation into macroeconomic models. The first is the balance sheet effect for financial intermediaries: In order to analyze how shocks in the financial sector spill over into the whole economy, neither the non-financial corporate balance sheet nor the integrated bank–firm balance sheet is satisfactory. The

\textsuperscript{11}Holding cash can be viewed as a stronger form of a holding option to wait.
second is the effect of imperfections in the interbank lending market: Properly modeling the interbank lending market and the interconnections between financial intermediaries is a prerequisite to modeling endogenous systemic risk.

Many authors of the earlier macroeconomics literature on financial frictions focus on credit market constraints on nonfinancial corporate borrowers without any real role for financial intermediation (e.g., Bernanke & Gertler 1989; Bernanke, Gertler & Gilchrist 1999; Kiyotaki & Moore 1997). Among them, Bernanke, Gertler & Gilchrist (1999) introduce a financial accelerator into the New Keynesian DSGE framework. However, they all focus on the balance sheet effect of nonfinancial firms while ignoring the unique properties of financial intermediaries. The financial accelerator channel from the balance sheet of nonfinancial corporate borrowers is definitely a relevant financial friction, but it is only one aspect of many possible financial frictions.

One of the first papers that tried to incorporate the financial sector into macroeconomic models and studied the effects of financial intermediary balance sheets and the interbank lending market is that by Gertler & Kiyotaki (2010). Others analyze monetary policies with liquidity risks in the interbank markets (e.g., Freixas, Martin & Skeie 2011). Gertler & Kiyotaki (2010) focus on understanding how disruptions in financial intermediation can induce a crisis that affects the real economy. The credit market constraints on financial intermediation are incorporated into an RBC framework, modified with habit formation and flow investment adjustment costs instead of capital adjustment costs. More precisely, the financial intermediaries are assumed to play three unique roles in the economy, as discussed extensively in the literature. First, the financial intermediaries are delegated monitors and specialists. They are conduits that channel funds from households to nonfinancial firms. Second, the financial intermediaries engage in maturity transformation. In the model, they are assumed to issue short-term debts and hold long-term assets. Third, the financial intermediaries facilitate liquidity provision via the interbank lending market. In the model, it is assumed that there is a continuum of “banks” that fund the goods producers and finance their investment from both a wholesale market, namely an interbank lending market, and a retail market, where banks hold deposits from households. To simplify this analysis, the authors assume constant returns to scale production, perfect labor mobility, and goods producers without financial constraints. With these assumptions, there is no need to keep track of the distribution of capital held by producers or their net worth. It is also assumed that the banks and the nonfinancial firms are “buddies,” in the sense that there is no financial friction in their funding relationship. In other words, it is essentially assumed that the producers’ balance sheet can be viewed as part of the banks’ balance sheet. To achieve such a simplification, following Gertler & Karadi (2011), the authors assume complete consumption insurance among workers and bankers, and independent and identically distributed random turnovers between workers and bankers. By doing so, this guarantees that a representative household exists to determine aggregate consumption and prices, with no need to track the wealth distribution of households. In this complete market economy, there is a unique SDF while the agents are actively borrowing and lending in equilibrium.

In contrast, He & Krishnamurthy (2013) incorporate the financial intermediary into a standard macrofinance model with segmentation between bankers and workers and reasonable terms of macroeconomic calibration in order to generate extremely nonlinear risk premium dynamics. They model the dynamics of risk premia during crises in asset markets where the marginal investor is a financial intermediary. In this model, intermediaries face an equity capital constraint. Risk premia rise when the constraint binds, reflecting the capital scarcity. The calibrated model matches the nonlinearity of risk premia during crises and the speed of reversion in risk premia from crisis levels back to precrisis levels. They evaluate the effect of three government policies: reducing intermediaries’ borrowing costs, injecting equity capital, and purchasing distressed assets. Injecting equity capital is particularly effective because it alleviates the equity capital
constraint that drives the model’s crisis. However, it is still far from satisfactory for monetary policy decision making because the model simplifies some important features; as a result, there is no way to see how the constrained financial intermediaries would affect the economy as a whole in such a model.

If the theory of He & Krishnamurthy (2013) is correct, the marginal value of wealth for financial intermediaries should provide a more informative SDF than that of a representative consumer. Empirically, Adrian, Moench & Shin (2010) use shocks to the leverage of securities broker-dealers to construct an intermediary SDF. Intuitively, deteriorating funding conditions are associated with deleveraging and a high marginal value of wealth. Their single-factor model prices size, book-to-market ratio, momentum, and bond portfolios with an $R^2$ of 77% and an average annual pricing error of 1%, performing as well as standard multifactor benchmarks designed to price these assets. They empirically document that financial intermediary balance sheets contain strong predictive power for future excess returns on a broad set of equity, corporate, and US Treasury bond portfolios. They also show that the same intermediary variables that predict excess returns forecast real economic activity and various measures of inflation. Their findings point to the importance of financing frictions in macroeconomic dynamics and provide quantitative guidance for preemptive macroprudential and monetary policies.

Moreover, Gilchrist & Zakrajsek (2012) empirically relate the predictive power of bond premia for the business cycle to the risk-bearing capacity of the marginal investors in these bonds. These investors act in a more risk-averse way when their capital becomes impaired, which translates to an increase of the bond premium and a reduction in the supply of credit available to potential borrowers.

However, the literature of occasionally binding financial constraints on financial intermediation (e.g., Brunnermeier & Sannikov 2014; Danielsson, Shin & Zigrand 2011; He & Krishnamurthy 2012) emphasizes that precautionary effects can generate endogenous tightening of margins. This is in contrast to the literature on the financial accelerator (e.g., Bernanke & Gertler 1989; Bernanke, Gertler & Gilchrist 1999; Christiano, Rostagno & Motto 2010; Gertler & Kiyotaki 2010; Kiyotaki & Moore 1997), in which the financial constraints are always binding. The precautionary effect means that even if the borrowing constraint is not currently binding, an increase in likelihood that it could be binding in the future, possibly due to increased uncertainty, can induce a tightening of margins.

The Christiano–Motto–Rostagno (CMR) framework, introduced by Christiano, Rostagno & Motto (2010), has been adopted by the ECB. However, it is still subject to some major concerns. First, while a crisis in the CMR framework can now originate in the financial sector, rather than through risk in the production sector, the interbank market is still missing. Second, the model poorly accounts for the external financing premium variables (e.g., the credit spread), negating the important advantage of modeling the financial sector separately from the real sectors. Third, the log-linearization method fails to capture key dynamics of the financial sector and the real economy. Fourth, the absence of precautionary effect makes it difficult to generate realistic nonlinear dynamics in asset prices.

### 3.7. Goods Markets and Markups

In the baseline New Keynesian DSGE model, the desired markup of price over marginal cost is constant. This is mainly due to two factors: a constant elasticity of substitution among differentiated goods and the validity of the Modigliani–Miller theorem. As observed by Blanchard (2009), however, the desired markup appears to be anything but constant, and how it varies in response to other factors is still unknown in macroeconomics.
One popular model for the desired markup is the customer market mechanism, in which a firm that lowers its current price not only sells more to its existing customers but also expands its customer base, leading to higher future sales at any given price. This idea was first introduced by Phelps & Winter (1970) and formalized by Gottfries (1986), Klemperer (1987), Farrell & Shapiro (1988), and Bils (1989), among others. Several strands in the marketing and industrial organization literature provide empirical support for the customer model. For example, Houthakker & Taylor (1966) use 80 detailed items of consumer expenditure and find that current demand depends positively on existing inventory, suggesting some habit formation. Guadagni & Little (1983) use a multinomial logit model of brand choice, calibrated on 32 weeks of purchases of regular ground coffee by 100 households, and show high statistical significance for the explanatory variables of brand loyalty. Erdem (1996) finds that for margarine, peanut butter, yogurt, and liquid detergent, accounting for habit formation improves both in-sample and out-of-sample fit. Genesove & Mullin (1997) find that the American Sugar Refining Company sharply cut its price to maintain market share and deter entry of competitors. Bronnenberg, Dhar & Dubé (2009) find an early entry effect on a brand’s current market share and perceived quality across US cities. Direct evidence has also been provided by firm-level surveys in several OECD countries (e.g., Amirault, Kwan & Wilkinson 2006; Aucremanne & Druant 2005; Fabiana et al. 2005; Hall, Walsh & Yates 1997), all pointing out that price stickiness is driven mainly by customer relationships. Dou & Ji (2020) build on the idea of the customer market and analyze how financing decisions interact with strategic pricing when the financial market is imperfect. One major focus of Dou & Ji (2020) is the endogenous relationship between financing and the price-setting behavior of firms. In a closely related paper, Chevalier & Scharfstein (1994) study the impact of imperfect financial markets on firms’ price-setting decisions. However, they use limited supermarket data to test the causal effect of liquidity shocks on goods price.

Chevalier & Scharfstein (1994) were the first to introduce capital market imperfection into a customer market model in an attempt to interpret countercyclical markups. They focus on how liquidity constraints affect pricing behavior and find that liquidity-constrained firms have an incentive to raise prices in order to boost current profits to meet their liabilities and finance investment. Gilchrist et al. (2017) provide more evidence using product-level price data. They find that, during the Financial Crisis of 2008, firms with weak balance sheets increased prices significantly relative to industry averages, whereas firms with strong balance sheets lowered prices. A general equilibrium model with financial market distortions is proposed to rationalize these findings. The idea of Dou & Ji (2020) is related to those of Chevalier & Scharfstein (1994) and Gilchrist et al. (2017), but Dou & Ji (2020) look for more evidence on firms’ pricing and financing behavior in normal times across industries and on the frequency of price resetting over business cycles. Their DSGE model provides a unified theory on Tobin’s q, corporate investment, financing, price setting, and asset pricing. Weber (2013) shows that firms that adjust their product prices infrequently earn a cross-sectional return premium of more than 4% a year.

The cyclicity of markups has significant implications for economic fluctuations, since countercyclical markups would tend to dampen fluctuations in economic activity, whereas procyclical markups would amplify fluctuations. To generate procyclical factor prices, Rotemberg & Woodford (1991) propose that markups should be countercyclical, and they provide evidence using aggregate data. The countercyclical markup is also found by Bils (1987) and in the supermarket industry (Chevalier & Scharfstein 1994). However, other studies find that markups are procyclical using different industry-level data (e.g., Domowitz, Hubbard & Petersen 1986; Ghosal 2000; Machin & Van Reenen 1993; Nekarda & Ramey 2013).

On the basis of the customer market model introduced by Phelps (1998), Gilchrist et al. (2017) investigate the effect of financial conditions on price-setting behavior during the Financial Crisis.
of 2008 by assuming a deep habit component in the model. In their model, firms have an incentive to set a low price to invest in market share. In other words, the loss from setting a lower price can be viewed as an investment cost for positive net present value projects (i.e., market shares). When financial distortions are severe, firms forgo these investment opportunities and maintain high prices because the marginal value for cash dominates the profits from investment in market share. The model with financial distortions implies a substantial attenuation of price dynamics in response to contractionary demand shocks relative to the baseline without financial distortions, which has important policy implications. Empirically, the authors find theory-consistent evidence that, at the peak of the crisis, firms with relatively weak balance sheets increased prices, while firms with strong balance sheets lowered prices.

3.8. Solution, Estimation, and Evaluation

The proper methodologies for the solution, estimation, and evaluation of New Keynesian DSGE models are critically important in economics, yet extremely difficult to do technically. Without proper methods, the credibility of monetary policies based on these models will be dramatically compromised, and their results may be extremely misleading, even if the modeler constructs a perfect model, one that incorporates all the mechanisms discussed in the previous sections. In this section, we review the methodologies used by central banks and highlight their principal issues. We emphasize that we do not advocate for the complexity of the DSGE model to keep increasing, even when the estimation of a super-nonlinear DSGE model can be done, because of the advanced computing techniques.

3.8.1. Solution methods. Solving the DSGE model with heterogeneous agents in incomplete markets and severe nonlinearity is mathematically equivalent to solving a large system of nonlinear equations. The nonlinearity and infinite dimensionality of the model make the problem extremely challenging, even for mathematicians and computer scientists. Given these technical and computational challenges, economists must make difficult trade-offs between complexity and tractability when specifying the model.

Because of these trade-offs, macroeconomists at central banks prefer to use simpler models and the log-linearization solution method. The DSGE model relies on log-linearization around the steady state. As pointed out by Tovar (2009), due to the computational burden often associated with the likelihood evaluation for the solution of the nonlinear expectation equations implied by DSGE models, the empirical literature has concentrated its attention on the estimation of first-order linearized DSGE models. First-order approximations have been, until recently, the main tool employed for empirically evaluating DSGE models and forecasting. However, as Judd (1997, p. 911) observes, “[i]f theoretical physicists insisted on using only closed-form solutions or proofs of theorems to study their models, they would spend their time examining the hydrogen atom, universes with one star, and other highly simplified cases and ignore most interesting applications of physical theories.”

The log-linearization approximation method has several important drawbacks. First, the solution methodology makes it impossible to model and study systemic risk. More recent papers on modeling financial intermediaries, such as those by Brunnermeier & Sannikov (2014) and He & Krishnamurthy (2013), show that the nonlinearity of the amplification effect is a key aspect of systemic risk. Second, first-order approximations are not appropriate for evaluating welfare across policies that do not affect the steady state of the economy, for instance, when asset prices and the risk premium are taken into consideration. Log-linearization around a constant steady state is not applicable to asset pricing because, by construction, it eliminates all risk premia in the
model. In fact, the risk premium is zero in a first-order approximation and constant in the case of a second-order approximation; therefore, higher-order approximations are required. Third, Fernández-Villaverde, Rubio-Ramírez & Santos (2006) consider log-linearization approximation to be unsatisfactory, arguing that second-order approximation errors in the solution of the model can have first-order effects on the likelihood function approximation. Ackerberg, Geweke & Hahn (2009) made important asymptotic corrections to a theoretical result from Fernández-Villaverde, Rubio-Ramírez & Santos (2006), arguing that the approximation error on the classical maximum likelihood estimation of the approximate likelihood function has the same magnitude as the approximation error of equilibrium policy functions. When exact yet highly nonlinear policy functions are approximated by local linear ones, the likelihood implied by the linearized model can diverge greatly from that implied by the exact model, and similarly the likelihood-based point estimation.

3.8.2. Estimation methods. Today, most central banks have adopted Bayesian likelihood estimation methods instead of the more traditional equation-by-equation estimation used for large macro models. The main reasons are as follows. First, as shown by Canova (2009), the likelihood function of DSGE models is often flat and irregular in a number of parameters. Prior information helps overcome such identification issues. (However, there are general issues regarding justifying the correct choice of priors, and it is dangerous to use too strong of a prior.) Second, the Bayesian approach can explicitly deal with measurement errors, unobservable state variables, large data sets, and different sources of information. Third, the Bayesian approach allows for decision making under uncertainty for policy makers. Fourth, although the Bayesian method is exposed to the “stochastic singular” problem that occurs when the number of variables is more than the number of shocks, there are some useful techniques to tackle the problem. This can be viewed as an example of the models lying between the data-driven and structural ends of the model spectrum.

The second main reason macroeconomists at central banks resort to log-linearization approximation is to make estimation easier. Since Smets & Wouters (2003), the Bayesian estimation method has become the most popular estimation approach at central banks. However, as is well known, the standard Bayesian method requires full specification of the likelihood function of the model. This seems implausible for complex New Keynesian DSGE models without log-linear approximation. However, advanced Bayesian computing techniques such as the approximate Bayesian computation (ABC) method can be adopted. This method is able to work with the Dynare software platform, which allows higher-order approximations of the model. Finally, when using these methods it is important to match the impulse response instead of matching only the moments of the model.

3.8.3. Evaluation methods. The traditional method of evaluating DSGE models is to compare the simulated subset of moments with those observed in the data. More cautious researchers have conducted sensitivity analyses to check the fragility of the model. However, they conduct this robustness check mostly in an informal way, by perturbing parameters one by one and measuring the difference in model effects. The parameter choices and magnitude of disturbance are ad

---

12 See, for example, Schmitt-Grohé & Uribe (2004), Kim et al. (2005), and An & Schorfheide (2007) for a discussion of second-order approximations.

13 For example, Harrison & Oomen (2010) added “structural shocks” to the baseline model to overcome the stochastic singular issue and improve the fitting of the data.

14 Chen, Dou & Kogan (2013) present a simple example using the ABC method to estimate a dynamic macroeconomic finance model.
hoc. Chen, Dou & Kogan (2013) observe that even when the model is stable in each parameter, it could be the case that the model is fragile in a combination of multiple parameters. Zin (2002) points out that a primary goal of characterizing asset market data using a tightly parameterized general equilibrium model is to try to uncover deep structural parameters for policy purposes. However, he emphasizes that it is not an easy task, mainly because the aggregate historical data are usually not enough to provide an informative statistical test on the models or their structural stability. To demonstrate this idea more explicitly, he discusses a simple asset pricing model that cannot be rejected by data if the asset pricing moment restrictions depend on high-order moments (e.g., the fifth moment) of the distribution of the fundamental process (i.e., the endowment process). It is difficult to visualize or even describe high-order moments of fundamental processes, so it is difficult to believe these asset pricing explanations would be deemed structural or useful. However, these high-order moments could be macroeconomic “dark matter” (as in Chen, Dou & Kogan 2013). The solution proposed by Zin (2002) is to augment the statistical tests with subjective non-sample-based judgments about the reasonableness of the assumptions. Chen, Dou & Kogan (2013) significantly improve Zin’s argument by explicitly defining and quantitatively measuring dark matter, while Zin (2002) demonstrates the idea only qualitatively. More importantly, Zin (2002) focuses only on the “weak identification” side of dark matter, while missing the more important side: that it may cause model implications to become extremely sensitive to parameters. While both Zin (2002) and Chen, Dou & Kogan (2013) emphasize the insufficiency of current statistical tests for structural model evaluation, Chen, Dou & Kogan (2013) propose an explicit, quantitative, and implementable method focused on dark matter or fragility in models to augment conventional statistical specification tests for model evaluation.

4. CONCLUSION

Just as the Great Depression and its aftermath inspired Tinbergen and Klein, and the recession and stagflation of the 1970s inspired Lucas, Kydland, and Prescott, the depth and length of the Financial Crisis of 2008 have given renewed urgency and relevance to macrofinancial economists around the world. Although this challenge may seem daunting, it is also an extraordinary opportunity to effect dramatic change in how we conduct macroeconomic policy. Three major themes have emerged in how we prioritize those changes.

The first theme is to take risk seriously in macroeconomic models and incorporate individual, institutional, and regulatory responses to changing risks, both actual and perceived. Thanks to early attempts to model the macroeconomy, the critical field of national income accounting emerged and transformed macroeconomics from armchair quarterbacking to a scientific endeavor with enormously practical implications. We now measure many aspects of the economy such as inflation, output, and unemployment, but have no measure of aggregate risk in the economy. The old adage that one cannot manage what one does not measure is particularly relevant when it comes to risk in the macroeconomy. In the same spirit of Keynes’s hope that economic policy would some day be as effective and prosaic as going to the dentist, we can hope that systemic risk measurement will some day be as effective as hurricane forecasts and flood warnings issued by the National Weather Service.

The second theme is to incorporate the intricacies of the financial sector more effectively into existing DSGE models. Given the complexity of today’s financial system, this challenge may seem hopeless and naive. However, the very essence of macroeconomics is to distill complex phenomena into macroscopic narratives that can be grasped and managed by human cognition. Together with new technologies such as massive data sets, new computational and statistical methods, and social media, the potential for creating even greater information compression for macroeconomic
policy decisions has never been more promising. The ability to measure business activity, leverage, inflation, and employment in real time at the level of the individual is close at hand, and the aggregation of such microlevel measures will surely transform macroeconomics.

The third theme, and perhaps the most radical, is to challenge the physics- and theory-based orthodoxy of macroeconomic modeling in reexamining the microfoundations of the DSGE framework. To bring models closer to reality, it may be necessary to let go of the deeply cherished conviction that agents always optimize their behavior according to rational expectations, and allow for certain predictable irrationalities in their behavior. These agents would still reflect the spirit of the Lucas critique by adapting to economic circumstance, but not necessarily in an instantaneously and fully optimal way. One of the positive aspects of financial crises may be to provide motivation for economists to revisit their assumptions of optimizing, forward-looking behavior and adjust them to reflect the realities of decision making in a complex, uncertain, and changing environment with limited information and cognitive abilities. This may also require us to abandon our predilection for simple models with elegant closed-form solutions in favor of less elegant but more practically relevant computational and numerical approaches to macroeconomic analysis.

When Albert Einstein was criticized for the complexity of his theory of relativity, he responded that “a theory should be as simple as possible, but no simpler.” The same can be said about the theories of the macroeconomy. We are discovering—as Keynes discovered over half a century ago—that from a policy perspective, being precisely wrong is not as useful as being approximately right.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Tobias Adrian, Richard Berner, Michael Binder, Marco Del Negro, Fernando Duarte, Mark Gertler, Lars Peter Hansen, Nobuhiro Kiyotaki, Nellie Liang, Debbie Lucas, Robert C. Merton, Johannes Pfeifer, David Skeie, Volker Wieland, and seminar participants at the Macro Financial Modeling Group 2013 Meeting at Chicago, Macro Financial Modeling Group 2015 Meeting at New York, and 1st Research Conference of the Center for Economic Policy Research Network on Macroeconomic Modeling and Model Comparison (MMCN) for helpful discussion and comments, and we thank Jayna Cummings for editorial assistance. We appreciate the generous support from the Macro Financial Modeling Group, the Alfred P. Sloan Foundation, and the MIT Laboratory for Financial Engineering.

LITERATURE CITED


Contents

Robert C. Merton: The First Financial Engineer
Andrew W. Lo ................................................................. 1

Robert C. Merton and the Science of Finance
Zvi Bodie ........................................................................ 19

The Information View of Financial Crises
Tri Vi Dang, Gary Gorton, and Bengt Holmström ......................... 39

Refinancing, Monetary Policy, and the Credit Cycle
Gene Amromin, Neil Bhutta, and Benjamin J. Keys ....................... 67

Macroeconomic Models for Monetary Policy: A Critical Review from a Finance Perspective
Winston W. Dou, Andrew W. Lo, Ameya Muley, and Harald Uhlig .......... 95

Global Banking: Toward an Assessment of Benefits and Costs
Claudia M. Buch and Linda S. Goldberg ..................................... 141

Credit Default Swaps: A Primer and Some Recent Trends
David Lando ........................................................................ 177

Debt Structure
Paolo Colla, Filippo Ippolito, and Kai Li ........................................ 193

Conflicts of Interest in Asset Management and Advising
Chester S. Spatt ..................................................................... 217

Strategic Decisions in Takeover Auctions: Recent Developments
B. Espen Eckbo, Andrey Malenko, and Karin S. Thorburn ................. 237

Portfolio Choice Over the Life Cycle: A Survey
Francisco Gomes .................................................................... 277

The Global Equilibrium Real Interest Rate: Concepts, Estimates, and Challenges
Michael T. Kiley .................................................................... 305
Informed Options Trading Before Corporate Events
Patrick Augustin and Marti G. Subrahmanyam ................................. 327

Textual Analysis in Finance
Tim Loughran and Bill McDonald ................................................. 357

Institutions and Innovation
Jie (Jack) He and Xuan Tian ....................................................... 377

Indexes
Cumulative Index of Contributing Authors, Volumes 5–12 .................... 399
Cumulative Index of Article Titles, Volumes 5–12 ............................. 402

Errata
An online log of corrections to Annual Review of Financial Economics articles may be found at http://www.annualreviews.org/errata/financial