

# Twenty Years of Time Series Econometrics in Ten Pictures

James H. Stock and Mark W. Watson

**T**wenty years ago, empirical macroeconomists shared some common understandings. One was that a dynamic causal effect—for example, the effect on output growth of the Federal Reserve increasing the federal funds rate—is properly conceived as the effect of a shock, that is, of an unanticipated autonomous change linked to a specific source. Following Sims (1980), the use of vector autoregressions to estimate the dynamic causal effect of shocks on economic variables was widespread. There was also an understanding that vector autoregressions, because they impose as little structure on the data as possible, cannot answer questions about changes in policy regimes, such as the macroeconomic consequences of the Fed adopting a new policy rule. For such questions, more structured models grounded in economic theory are needed. At the same time, there was an increasing recognition that the available methods needed significant work. The schemes used to identify structural shocks in vector autoregressions were often seen as unconvincing by researchers outside the field, and the small structural models of the time were not econometrically estimated, miring that enterprise in an unhelpful debate over how to calibrate such models. In addition, there were chinks emerging in the theoretical econometric underpinnings of inference in time series data, as well as opportunities for using the much larger datasets becoming available, if only the tools to do so could be developed. The time was ripe for progress.

■ *James H. Stock is the Harold Hitchings Burbank Professor of Political Economy, Harvard University, Cambridge, Massachusetts. Mark W. Watson is Howard Harrison and Gabrielle Snyder Beck Professor of Economics and Public Affairs, Princeton University, Princeton, New Jersey. Their email addresses are james\_stock@harvard.edu and mwatson@princeton.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.31.2.59>

doi=10.1257/jep.31.2.59

This review tells the story of the past 20 years of time series econometrics through ten pictures. These pictures illustrate six broad areas of progress in time series econometrics: estimation of dynamic causal effects; estimation of dynamic structural models with optimizing agents (specifically, dynamic stochastic equilibrium models); methods for exploiting information in “big data” that are specialized to economic time series; improved methods for forecasting and for monitoring the economy; tools for modeling time variation in economic relationships; and improved methods for statistical inference.

These pictures remind us that time series methods remain essential for shouldering real-world responsibilities. The world of business, finance, and government needs reliable information on where the economy is and where it is headed. Policy-makers need analysis of possible policies, and macroeconomists need to improve their understanding of the workings of modern, evolving economies. Taken together, the pictures show how 20 years of research have improved our ability to undertake these professional responsibilities. These pictures also remind us of the close connection between econometric theory and the empirical problems that motivate the theory, and of how the best econometric theory tends to arise from practical empirical problems.

A review of 20 years of research must make some arbitrary decisions. One of our decisions is to focus on empirical macroeconomics, not finance. Fortunately, there are good surveys of the many developments in financial econometrics: for example, see the papers in Aït-Sahalia and Hansen (2010). Another concerns the choice of figures. Our ten figures are not meant to single out superstar papers (although some are) but rather to represent important lines of research: each figure illustrates a broader research program. In choosing these figures, we first looked for influential early papers from the late 1990s and early 2000s that framed subsequent research. This yielded five figures from papers with an average of 1,486 Google Scholar citations each. We then looked for figures more recently published that illustrate key findings or methods in a relatively mature line of research, yielding four more figures. Our final figure, which is not from published research, illustrates an open empirical challenge for research ahead.

## **Causal Inference and Structural Vector Autoregressions**

An ongoing question in empirical macroeconomics is how to determine the causal effect of a policy change. For example, what is the effect of an autonomous, unexpected, policy-induced change in the monetary policy target rate—that is, a monetary policy shock—on output, prices, and other macro variables? The underlying problem is simultaneous causality: for example, the federal funds interest rate depends on changes in real GDP through a monetary policy rule (formal or informal), and GDP depends on the federal funds interest rate through induced changes in investment, consumption, and other variables. Thus, one cannot determine the effect of a change in the federal funds interest rate simply by using the rate (perhaps along with lagged values of the rate) as a right-hand-side variable in a regression to explain

GDP. Somehow, a researcher needs to isolate the exogenous variation in the federal funds interest rate, and for that you need external information.

Since the seminal work of Sims (1980), vector autoregressions have been a standard tool for estimating the causal effects over time of a shock on a given macro variable. This tool evolved into “structural” vector autoregressions, which are based on the idea that the unanticipated movements in the variables—that is, their forecast errors—are induced by structural shocks. The goal of structural vector autoregressions is to impose sufficient restrictions so that one or more structural shocks can be identified: specifically, that one or more shocks can be represented as an estimable linear combination of the forecast errors. The result of this analysis is the estimation of a dynamic path of causal effects, which in macroeconometrics is called a “structural impulse response function.”

However, many applications of the original methods for identification of structural autoregressions that were dominant in the 1980s and 1990s have not withstood close scrutiny (as articulated, for example, by Rudebusch 1998). For example, a popular method for identifying monetary policy shocks in the 1980s and 1990s was to assume that economic activity and prices respond to a monetary policy shock with a lag, but that monetary policy responds systematically to contemporaneous nonmonetary shocks to the other variables. Under this assumption, the predicted value in a regression of the federal funds rate on its lags and on current and lagged values of the other variables is the endogenous policy response, and the residual is the unanticipated exogenous component—that is, the monetary policy shock. But this identifying assumption is not credible if the other variables include other asset prices, such as long-term interest rates.

Thus, this area needed new approaches. Broadly speaking, these new approaches bring to bear external information: information outside the linear system of equations that constitutes the vector autoregression. The development of new methods for estimating causal effects has been one of the main advances in microeconometrics over the past two decades (as discussed in several other articles in this symposium), and the focus on credible identification has parallels in the structural vector autoregression literature.

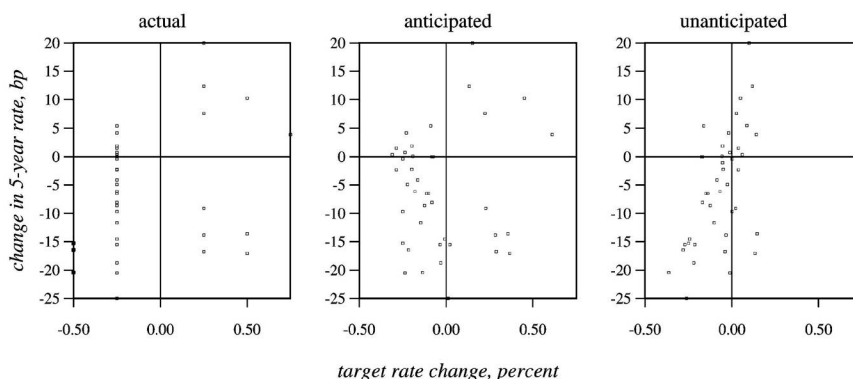
### **Using External Information to Estimate the Shock Directly**

This brings us to our first picture, which is from Kuttner (2001). Kuttner’s interest was in estimating the dynamic causal effect of a monetary policy shock on long-term interest rates, which is part of the broader program of estimating their dynamic causal effect on macroeconomic variables. Because the Fed controls the federal funds interest rate, one might initially think that the fed funds rate is exogenous; but not so, because some of the changes are responses to changes in economic activity which have their own effect on long-term interest rates. Rather, the exogenous part of the fed funds rate—the monetary policy shock—is the part that is not a response to economic activity. Kuttner’s innovation was to draw on external information to identify the shock. Specifically, he knew that the Federal Reserve Open Market Committee announced its decisions at a specific time after its meetings, and he also had evidence (along with the theory of efficient financial markets) that the

Figure 1

**Changes in the 5-year Treasury Rate and in the Target Federal Funds Rate on Federal Reserve Open Market Committee (FOMC) Announcement Dates**

(*fed funds changes are decomposed into anticipated and unanticipated components using changes in the fed funds futures market on announcement dates*)



Source: Kuttner (2001), Figure 2.

Note: The figure shows changes in the 5-year Treasury rate (on the y-axes) and in the fed funds target rate (on the x-axes) on Federal Reserve Open Market Committee (FOMC) announcement dates. Unanticipated changes in the fed funds rate—which are the monetary policy shocks—are identified as changes in the fed funds futures rate from before to after the announcement of a change in the FOMC’s target for the fed funds rate. The anticipated change is the actual change in the fed funds futures rate, minus the unanticipated change.

fed funds future rate was an efficient forecast of future fed funds rates. Thus, he was able to measure the unexpected part of the change in the federal funds futures rate as the change in the fed funds rate before and after the announcement. Assuming that no other relevant news was released during the announcement window, this change in the fed funds futures rate measures the change in market expectations of the fed funds rate resulting from the announcement—that is, it measures the monetary policy shock associated with the announcement. By using this external information, he could directly estimate the monetary policy shock.

Kuttner’s figure (our Figure 1) shows that this unanticipated component of the change in the target rate is associated with changes in the five-year Treasury rate (right panel), but anticipated changes are not (center). As a result, there is no particular relationship between the actual announced target and the five-year rate (left). We interpret this figure as a compelling plot of the “first stage” in instrumental variables regression: it shows that an instrument (the unanticipated component of the target change on the announcement day) is correlated with an endogenous variable (the five-year interest rate).

The idea of using external information to identify shocks for structural vector autoregression analysis traces back to Romer and Romer (1989), who used textual and historical information to identify some exogenous monetary policy shocks. In addition to Kuttner (2001), Cochrane and Piazzesi (2002), and Faust, Rogers,

Swanson, and Wright (2003), and Bernanke and Kuttner (2005) are early papers that use interest rate changes around Federal Reserve announcement dates to identify monetary policy shocks. In a similar spirit, Hamilton (2003) and Kilian (2008) use external information on international oil supply disruptions to estimate the effect of oil supply shocks on the economy.

This line of attack aims to measure the exogenous shock directly from external information, such as knowledge of the interest rate markets around announcement dates. If the shock can actually be measured, then estimation of structural impulse response functions is straightforward: because the shock is uncorrelated with other shocks, one can simply regress a variable of interest on current and lagged values of the shock, and the resulting coefficients trace out the dynamic causal effect (for example, Stock and Watson 2011, chap. 15). But doing so requires a particular strong form of external information: that the shock can be accurately measured.

### **Identification by External Instruments**

If the external information succeeds in measuring only part of the shock or produces a noisy measurement of the shock, then the measured shock has the interpretation as an instrumental variable and regressions on the measure have the interpretation as the first stage in two-stage least squares. Arguably, many of the shock measures proposed to date yield imperfect measures. For example, changes in federal funds futures around an announcement reveal only a part of the monetary policy shock. In this case, the external shock measure is an instrumental variable: it is exogenous (that is, it is uncorrelated with other structural shocks) if properly constructed, and it is relevant because it is correlated with the true shock. Hamilton (2003) uses his measured international oil shock measure as instrument in a single-equation setting. In a vector autoregression, the technicalities differ from standard instrumental variables regression because the observed endogenous variables are forecast errors, not the original variables themselves. Still, the two criteria for a valid instrument, relevance and exogeneity, are the same in the structural vector autoregression application as in standard instrumental variable regression.

The explicit use of external instruments in structural vector autoregressions is fairly recent. This method is described in Stock (2008), Ramey (2016), and Stock and Watson (2016). Empirical applications of identification of structural impulse response functions using external instruments include Stock and Watson (2012a), Mertens and Ravn (2013), and Gertler and Karadi (2015).

### **Identification by Heteroskedasticity**

Another method for identifying impulse response functions developed during the past 20 years exploits the observation that changes in the *variance* of the shocks can serve to identify the impulse response functions if those responses remain constant despite the heteroskedasticity of the shocks. Suppose that there are two known regimes, a high- and a low-volatility regime. Identification by heteroskedasticity works by generating two sets of moment equations, one for each regime. Although

neither set can be solved on its own (the identification problem), assuming that the impulse response functions are the same across both regimes imposes enough parametric restrictions that together the two sets of equations can be solved, and thus the impulse response functions can be identified. This clever insight was developed for regime-shift heteroskedasticity by Rigobon (2003) and Rigobon and Sack (2003, 2004), and for conditional heteroskedasticity by Sentana and Fiorentini (2001) and Normandin and Phaneuf (2004). Lütkepohl (2013) offers a survey and discussion.

### **Identification by Sign Restrictions**

An altogether different approach to identification in structural vector autoregressions is to use restrictions on the sign of impulse responses to identify the economic shocks. For many shocks, disparate macro theories often agree on the signs of their effects, at least over short horizons. Although several early papers build off this insight, the method developed by Uhlig (2005) is the most widely used. In his application, Uhlig restricted the impulse response with respect to a monetary policy shock identified by requiring that, on impact and over the next five months, the response of overall prices, commodity prices, and nonborrowed reserves to a contractionary monetary policy shock are not positive, and that the response of the federal funds interest rate is not negative. Identification using sign restrictions can be compelling and has been widely adopted.

At a mathematical level, using sign restrictions is fundamentally different than the other methods that identify shocks: with enough restrictions, those methods lead, in large samples, to a unique impulse response function, whereas the sign restrictions approach only determines a set that includes the impulse response. That is, sign-identified impulse response functions are not point-identified, but instead are set-identified.

Set identification of impulse response functions raises subtle issues of inference, which have only recently been appreciated. Following Uhlig (2005), the standard approach is Bayesian, but just as the identification scheme in classical structural vector autoregression methods can strongly influence results, the prior distribution over the unidentified region of the impulse response parameter space strongly influences Bayesian inference, even in large samples. These methods therefore require great care to produce transparent, valid, and robust inference. Recent papers tackling inference in sign-identified structural vector autoregressions are Fry and Pagan (2011), Moon, Schorfheide, and Granziera (2013), Giacomini and Kitagawa (2014), Baumeister and Hamilton (2015), and Plagborg-Møller (2016). For additional discussion and references to the recent methodological literature see Stock and Watson (2016, Section 4).

### **Estimation of Dynamic Stochastic General Equilibrium Models**

Dynamic stochastic general equilibrium models are models of forward-looking, optimizing economic agents who live in an economy subject to unexpected shocks. The development of methods for solving and estimating these models, combined

with their grounding in optimizing economic theory, has made them a central tool of monetary policy analysis at central banks.

One of the first full-system estimations of a dynamic stochastic general equilibrium model was by Ireland (1997), who estimated a three-equation (GDP, prices, and money) system by maximum likelihood. However, maximizing the likelihood proves far more difficult numerically than averaging over the likelihood using a Bayesian prior, and today the dominant methods for estimating dynamic stochastic general equilibrium models are Bayesian. These methods were first used by DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), and Otrok (2001) for small dynamic stochastic general equilibrium systems. Smets and Wouters (2003) showed that these methods can be applied to larger dynamic stochastic general equilibrium models that are rich enough to be a starting point for monetary policy analysis.

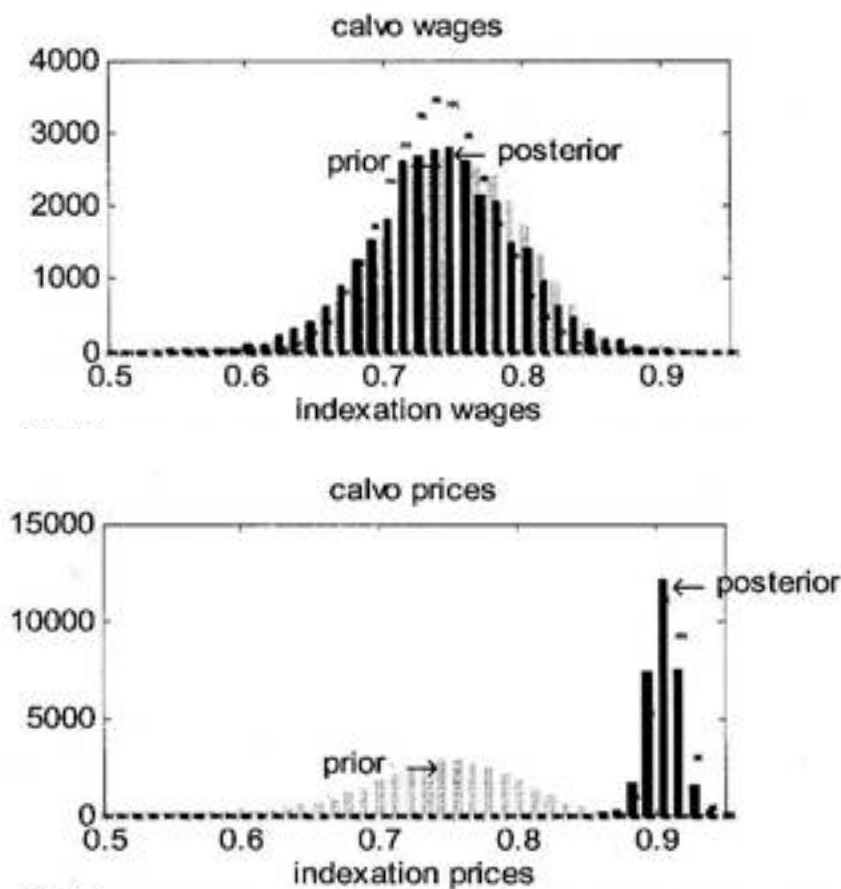
Figure 2, taken from Smets and Wouters (2003), represents the breakthroughs made over the past 20 years in the estimation of dynamic stochastic general equilibrium models. In their model, the “Calvo wage” parameter in the first panel is the probability that a worker’s wage does not change, and the “Calvo price” parameter in the second panel is the probability that the firm’s price does not change. As Figure 2 illustrates, the method works: The computational problems encountered when fitting dynamic stochastic general equilibrium models using frequentist methods such as maximum likelihood are sidestepped by computing posteriors, facilitated by a suite of tools developed in the modern Bayesian computational literature. For some parameters, such as the “Calvo price” parameter, the data are highly informative: incorporating the data results in much stickier prices than the authors’ prior, so that the posterior and prior distributions are quite different. But for other parameters, such as the “Calvo wage” parameter, the data are much less informative, so that the prior and posterior essentially coincide. Thus, the Calvo wage parameter is in effect calibrated by the researcher, so the resulting complete model combines estimation where the data are informative with calibration where they are not.

This property of estimation *cum* calibration means that care needs to be taken in interpreting measures of uncertainty arising from the model. From a frequentist perspective, a classic justification of Bayesian methods is that coverage intervals (“Bayes credible sets”) computed using the Bayesian posterior are essentially the same as frequentist confidence intervals in large samples, as long as a continuous prior does not rule out parameter values. (This is the celebrated Bernstein–von Mises theorem.) But for dynamic stochastic general equilibrium models, because the data are uninformative for some parameters—that is, some parameters are poorly identified—this equivalence does not hold and the uncertainty measures are heavily influenced by the shape of the prior. We return to this issue below, when we discuss weak identification.

The literature on estimation of dynamic stochastic general equilibrium models is vast and, because it quickly gets into specialized computational devices, it can be difficult to penetrate. For example, models of the Smets–Wouters sort rely on log-linearized approximations to decision rules, which both makes the models fairly easy to solve and means that the Kalman filter can be used to compute the Gaussian likelihood. Much of the recent methodological research on estimation of these models has focused on avoiding the log-linearization step. Among other things,



Figure 2

**Prior and Posterior Distributions for Two Structural Parameters in a Dynamic Stochastic General Equilibrium Model**

Source: Smets-Wouters (2003), Figure 1c (upper panel).

Note: This figures represents the breakthroughs made over the past 20 years in the estimation of dynamic stochastic general equilibrium models. In the model of Smets-Wouters (2003), the “Calvo wage” parameter in the first panel is the probability that a worker’s wage does not change, and the “Calvo price” parameter in the second panel is the probability that the firm’s price does not change.

avoiding log-linearization can improve the ability to analyze the effects of risk and uncertainty. However, there are substantial computational challenges in estimating nonlinear models, so that log-linearization remains common in practice. Canova (2007) provides an accessible textbook treatment of the linearize/Kalman filter/Bayes approach. Herbst and Schorfheide (2015) provide an up-to-date textbook treatment that focuses on computationally efficient methods for evaluating the posterior of linearized models. Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016) provide a detailed overview of methods that avoid linearization.



## Dynamic Factor Models and “Big Data”

The idea of using a large number of series to understand macroeconomic fluctuations is an old one, dating back at least as far as the economic indexes and forecasts of the Harvard Economic Service in the 1920s (Friedman 2009) and to Burns and Mitchell’s (1946) use of 1,277 time series to study business cycles. The challenge of using large numbers of series is the proliferation of parameters in standard time series models. While there were large macroeconomic models developed in the 1960s, and versions of them remain in use today, the restrictions that reduced the number of parameters in those models were heavily criticized as being arbitrary, having neither statistical nor economic foundations. Although low-dimensional vector autoregressions had become a standard macroeconometric tool by the mid-1990s, an outstanding challenge was increasing the number of variables, both to improve forecasting and to span a wider range of forecast errors, and thus structural shocks. The technical challenge was that in an unrestricted vector autoregression, the number of parameters increases with the square of the number of variables. Methods were needed to manage this proliferation of parameters if time series methods were to be used with large numbers of variables.

Dynamic factor models impose parametric restrictions in a way that is consistent with empirical evidence and a broad set of modern theoretical models. In a dynamic factor model, a given observable variable—say, the growth rate of consumption of services—is written as the sum of a common component and an idiosyncratic component. The common component depends on unobserved (or latent) common variables, called factors, which evolve over time; the idiosyncratic component is uncorrelated with the common component and has limited correlation with the other idiosyncratic components. The idiosyncratic component captures measurement error and series-specific disturbances that have no broader macroeconomic consequences. Thus, in a dynamic factor model, a small number of unobserved factors explain the comovements of a large number of macroeconomic variables.

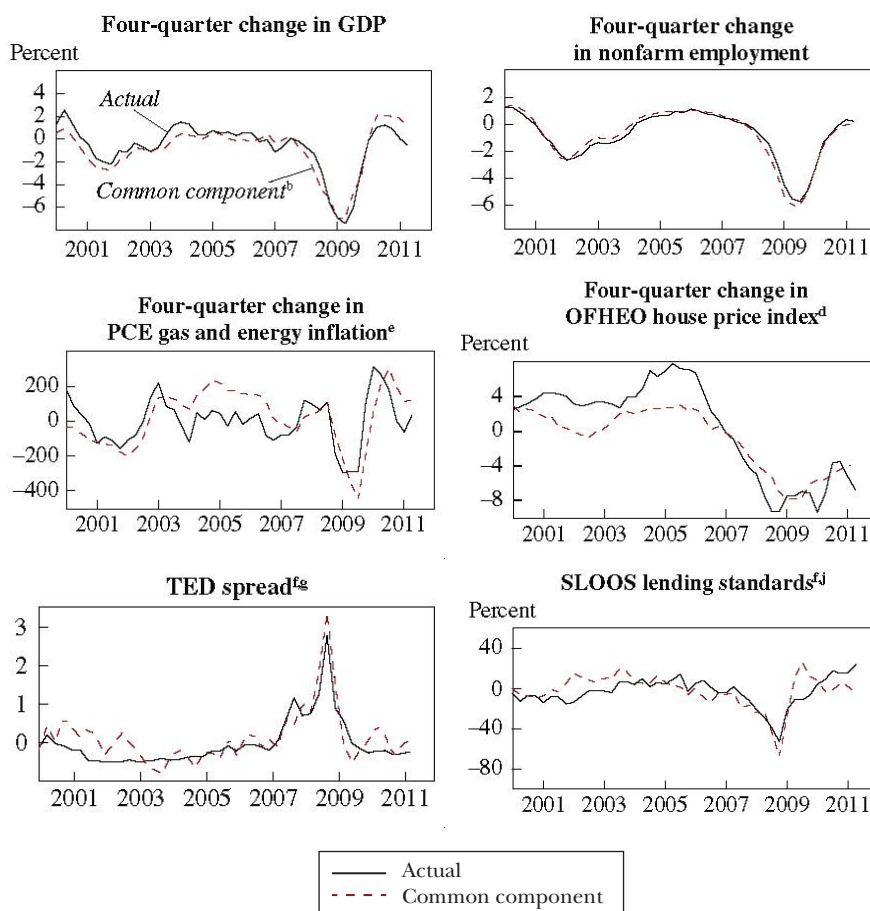
This brings us to our next figure, which is taken from Stock and Watson (2012a). Figure 3 shows the predicted value of six US quarterly macro variables from a 200-variable, six-factor dynamic factor model; this predicted value is called the “common component” of the series. The in-sample  $R^2$  of the common component for four-quarter growth in GDP (that is, the  $R^2$  of the regression of the four-quarter growth in GDP on the four-quarter growth of the six factors) is 73 percent; the average  $R^2$  of the common component over 21 major expenditures variables from the national income and product accounts is 56 percent; and the average  $R^2$  for all 200 variables is 46 percent. The parameters in this dynamic factor model were fitted using data from 1959–2007, so the post-2007 values of the common component represent the pseudo out-of-sample fit. At the visual level, for these and many other series, the fit is essentially the same in-sample and out-of-sample, suggesting that the parameters of the dynamic factor model remained largely stable during and after the financial crisis.

As Figure 3 illustrates, dynamic factor models fit the data. Techniques for dynamic factor analysis now can handle arbitrarily many series. One convenient way to estimate

Figure 3

**Selected US macroeconomic Time Series: Actual Values and Common Components**

(where the common components are the fitted values using the factors from a 200-variable, 6-factor dynamic factor model fit using data from 1959–2007)



Source: Stock-Watson (2012a), Figure 2.

Note: Figure 3 shows the predicted value of six US quarterly macro variables from a 200-variable, 6-factor dynamic factor model; this predicted value is called the “common component” of the series. The parameters in this dynamic factor model were fitted using data from 1959–2007, so the post-2007 values of the common component represent the pseudo out-of-sample fit.

the factors is principal components analysis, in which the factors are estimated by least squares. When estimated using many series, the principal component factor estimates can be treated as data for subsequent regressions (Stock and Watson 2002; Bai 2003; Bai and Ng 2006). To implement this approach, one needs to decide how many factors to use, and Bai and Ng (2002) show how to use information criteria to estimate the number of factors. This approach can be expanded to arbitrarily many series without substantially increasing the computational burden, indeed these models provide a twist on the usual “curse of dimensionality.” In dynamic factor models, the precision

of the estimation of the factors improves as the number of data series increases, so that the curse becomes a blessing.

Because of theoretical and empirical work over the past 20 years, dynamic factor models have become a leading method for the joint modeling of large numbers—hundreds—of economic time series. Dynamic factor models have natural applications to macroeconomic monitoring and forecasting, a topic we take up below. They also can be used to estimate the effect of a structural shock, such as a monetary policy shock, on multiple economic variables. These economy-wide shocks drive the common factors, and because the factors can be estimated, the economic shocks can be estimated up to a nonsingular linear transformation. As a result, the techniques for shock analysis developed for structural vector autoregressions, including the new methods discussed above, carry over directly to dynamic factor models. By using many variables, dynamic factor models can more plausibly capture macro-structural shocks than can low-dimensional vector autoregressions. Moreover, the estimated structural impulse response functions are internally consistent across all the variables. In Stock and Watson (2016), we survey dynamic factor models, with a focus on structural shock analysis.<sup>1</sup>

Dynamic factor models are not the only method available for high-dimensional modeling. A different approach is to use a Bayesian prior distribution over the vector autoregression parameters to reduce the influence of the data on any one parameter estimate and thus to reduce the amount of noise across parameter estimates. In some applications, large numbers of restrictions arise naturally: for example, global vector autoregression reduces the dimensionality of the vector autoregression parameter space by restricting domestic variables to depend on foreign variables only through a small number of weighted averages of global variables (Chudik and Pesaran 2016).

While this discussion has focused on the development of econometric methods for analyzing high-dimensional time series models, the other major development that has facilitated this work is the ready availability of data. The Federal Reserve Bank of St. Louis's FRED database, which migrated to an online platform in 1995, has been a boon to researchers and to the general public alike. A recent useful addition to FRED is FRED-MD, a monthly dataset currently comprised of 128 major economic time series for use in high-dimensional macroeconomic modeling (McCracken and Ng 2016); a beta-version with quarterly data (FRED-QD) is now available too. These datasets provide a common testbed for high-dimensional time series modeling and relieve researchers from the arduous task of updating a large dataset in response to new and revised data. A more specialized database, maintained by the Federal Reserve Bank of Philadelphia, archives and organizes

<sup>1</sup>A variant of a dynamic factor model is the factor-augmented vector autoregression (Bernanke, Boivin, and Elias 2005), in which one or more of the factors are modeled as observed. For example, because the Federal Reserve controls the federal funds interest rate, Bernanke, Boivin, and Elias (2005) argue that the target interest rate is itself a macroeconomic factor. Alternatively, factor-augmented vector autoregression can be interpreted as augmenting a low-dimensional vector autoregression with information from a first-step dynamic factor model. See Stock and Watson (2016) for a discussion of the relation between dynamic factor models and factor-augmented vector autoregressions.

real-time economic data; these data are especially valuable to those who want to test tools for real-time monitoring and forecasting.

## **Macroeconomic Monitoring and Forecasting**

Two important related functions of macroeconomists in business and government are tracking the state of the economy and predicting where the economy is headed. During the 1960s and 1970s, these two functions—macroeconomic monitoring and macroeconomic forecasting—relied heavily on expert judgment. The 1980s and 1990s saw new efforts by time series econometricians to place macroeconomic monitoring and forecasting on a more scientific footing: that is, to be replicable, to use methods that are transparent and have well-understood properties, to quantify uncertainty, and to evaluate performance using out-of-sample experience. While these advances provided macroeconomic monitoring and forecasting with a solid foundation, much work remained to be done. This work included improving methods for quantifying and conveying forecast uncertainty; dramatically expanding the number of data series that could be used, both to enable real-time monitoring to use the most recently released information and to improve forecasts; and developing reliable forecasting tools that take into account the evolution of the economy. Here, we discuss the first two of these: forecast uncertainty and macroeconomic monitoring. Issues of model instability go far beyond macroeconomic monitoring and forecasting, so we defer that discussion to the next section.

### **Estimating and Conveying Forecast Uncertainty**

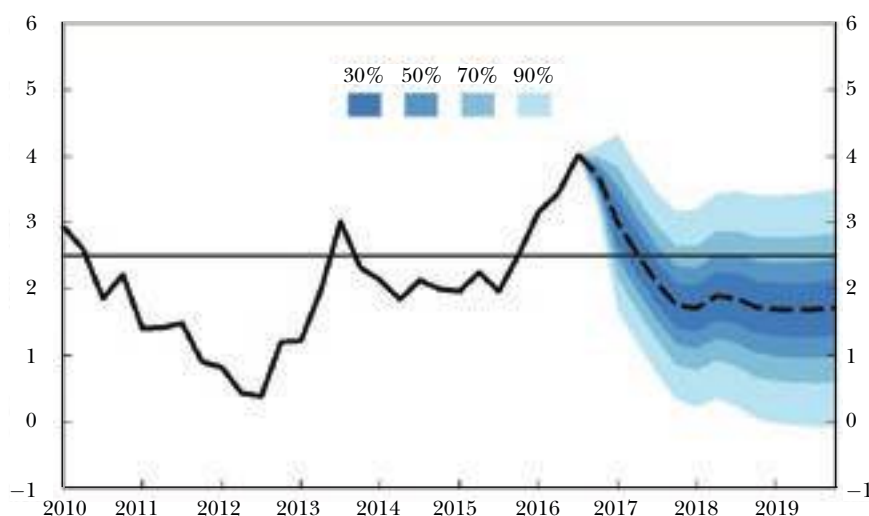
A fundamental problem of economic forecasting is that many economic variables are inherently very difficult to forecast, and despite advances in data availability, theory, and computational power, we have not seen dramatic improvements in forecast accuracy over the past decades. One implication of this observation is that economic forecasters should focus on communicating not just point estimates, but likely future ranges or distributions of the variable.

Our next figure highlights the development and adoption of density forecasts over the past 20 years. Figure 4 is a real-time release of a so-called fan chart from the Bank of Norway's Monetary Policy Report for December 2016. A fan chart communicates uncertainty by providing a density that describes the distribution of possible future values of the series being forecast, in this case Norwegian consumer price inflation.<sup>2</sup> The Bank of England was an early leader in the use of density forecasts and fan charts to communicate uncertainty to the public, and these methods are now widely adopted. Methods for constructing density forecasts are reviewed in

<sup>2</sup>The forecast uncertainty is better communicated in color! See the real thing at the websites of the Bank of England Inflation Report (<http://www.bankofengland.co.uk/publications/Pages/inflationreport>) and the Norges Bank Monetary Policy Report (<http://www.norges-bank.no/en/Published/Publications/Monetary-Policy-Report-with-financial-stability-assessment/>).

Figure 4

**Fan Chart (Density Forecast) for Consumer Price Index (CPI) inflation in Norway**  
(percent; four-quarter change)



*Source:* Reproduced from Chart 2.2c of Norges Bank Monetary Policy Report for December 2016, which used data from Statistics Norway and Norges Bank.

*Notes:* This chart shows the distribution of possible future values of Norwegian consumer price inflation, projections for 2016 Q4 through 2019 Q4.

Elliott and Timmermann (2016, ch. 13), and Corradi and Swanson (2006) survey methods for evaluating the accuracy of density forecasts.

Beyond the clear communication of uncertainty, the past 20 years of academic work on forecasting has focused on extending the scientific foundations for forecasting. These include methods for evaluating forecasts (including density forecasts), selecting variables for forecasting, and detecting forecast breakdown. While judgment will inevitably play a role in interpreting model-based forecasts, a central goal of this research program is to reduce the amount of judgment involved in constructing a forecast by developing reliable models and tools for evaluating those models. For a graduate textbook treatment, see Elliott and Timmerman (2016), and for additional detail see Elliott and Timmermann (2013).

### Macroeconomic Monitoring

Twenty years ago, economists who monitored the economy in real time used indexes of economic indicators and regression models for updating expectations of individual releases (such as the monthly employment report), combined with a large dose of judgment based on a narrative of where the economy was headed. While this approach uses data, it is not scientific in the sense of being replicable, using well-understood methods, quantifying uncertainty, or being amenable to later evaluation. Moreover, this method runs the risk of putting too much weight on the most recent but noisy data releases, putting too little weight on other data,

and being internally inconsistent because each series is handled separately. Because knowing the current state of the economy in real-time is an ongoing, arguably increasingly important responsibility of policymakers, time series econometricians at central banks and in academia have put considerable effort into improving the foundations and reliability of real-time macroeconomic monitoring.

Our next figure illustrates a central line of research in macroeconomic monitoring: the use of large models, in particular dynamic factor models, to incorporate real-time data releases to provide an internally consistent framework for estimating current economic conditions. Figure 5 is taken from the February 10, 2017, weekly update published by the New York Federal Reserve Bank. The dynamic factor model used by the New York Fed incorporates the most recently available data on 36 major economic indicators to provide a weekly estimate of the growth of GDP in the current quarter. Figure 5 shows the evolution of this real-time forecast of current-quarter GDP growth—for obvious reasons, called a “nowcast” of GDP—for the fourth quarter of 2016.

In August, the prognosis was for growth slightly above 2 percent at an annual rate, but by the first Friday in the fourth quarter (October 7), the nowcast had fallen to 1.3 percent. The November 18 nowcast rose to 2.4 percent on the strength of retail sales and housing starts data released that week. Then weak industrial production data, along with weak housing data released less than two hours before the December 16 update, pushed that nowcast down to 1.8 percent. As it happened, the advance estimate of fourth-quarter GDP growth released January 27 was 1.9 percent, slightly less than the estimate of 2.1 percent made on January 20.

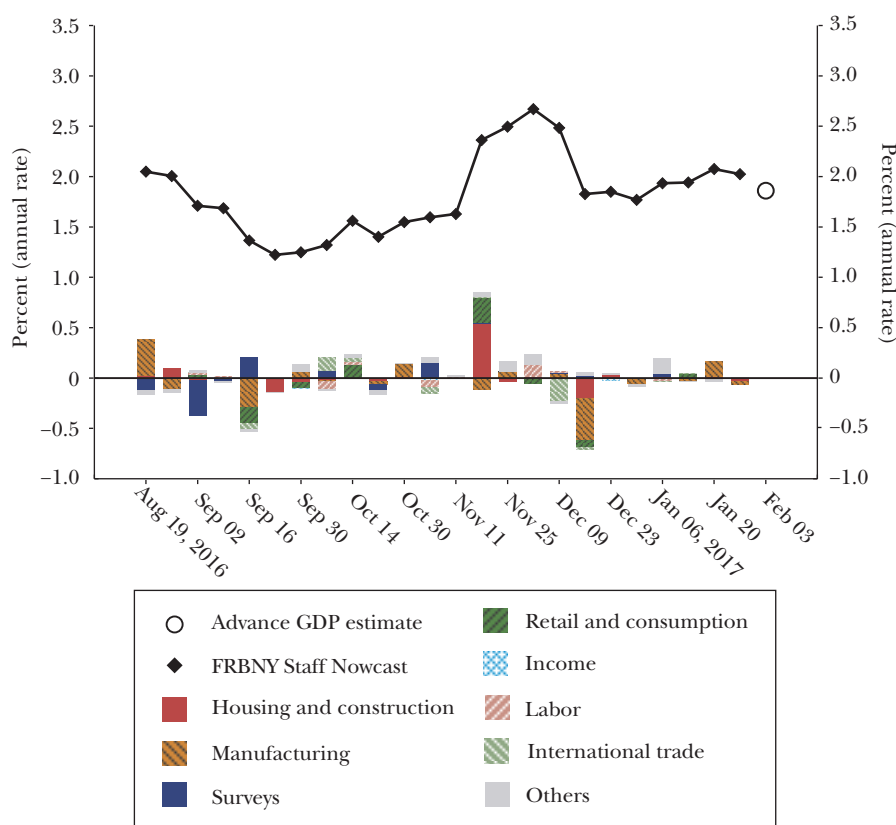
Under the hood of this real-time tracking product is a powerful set of tools for updating estimated factors in dynamic factor models using real-time data flows. The use of dynamic factor models for real-time macroeconomic monitoring incorporating staggered data releases dates to the NBER experimental coincident index (Stock and Watson 1989). By today’s standards, that index was primitive: a monthly release that encompassed only four variables. The current suite of tools for handling large series and complicated data flows are explicated in detail in Bańbura, Giannone, Modugno, and Reichlin (2013). The New York Fed’s model is updated (using the Kalman filter) as new data arrives, yielding an updated estimate of the single latent factor which in turn provides an updated estimate of the current-quarter value of GDP growth. By using a single flexible model, the news content of each series is exploited in a disciplined and internally consistent way. Some announcements contain substantial news, but many do not, and using a single model to evaluate these releases—rather than a suite of small models or judgment—provides a scientific way to use the real-time data flow.

The New York Fed report is one of several that use dynamic factor models to provide real-time, publicly available reports on the state of the economy. The EUROCOIN index, maintained by the Centre for Economic Policy Research and the Bank of Italy, is a real-time monthly index computed using a dynamic factor model with approximately 145 variables, calibrated to estimate monthly euro-zone GDP growth (Altissimo, Cristadoro, Forni, Lippi, and Veronese 2010). The Chicago Fed National Activity Index is a monthly index of real economic activity

Figure 5

### Contributions of Daily Data Releases to the Federal Reserve Bank of New York Real-Time Nowcast of 2016Q4 GDP Growth

(bars represent weekly contributions of data revisions to changes in the nowcast)



Source: Federal Reserve Bank of New York Nowcasting report, February 10, 2017.

Note: Figure 5 shows the evolution of a real-time forecast of 2016 fourth-quarter GDP—for obvious reasons, sometimes called a “nowcast.” Technically, the points through September 31, 2016, are forecasts of fourth quarter GDP growth; the points October 1 through December 31, 2016, are nowcasts; and the points January 1, 2017, to the end of the series are backcasts of fourth quarter GDP growth.

constructed as the single factor in an 85-variable dynamic factor model. The Federal Reserve Bank of Philadelphia maintains the Aruoba-Diebold-Scotti (2009) index, which is updated weekly using a six-variable dynamic factor model with one quarterly series (GDP), four monthly series, and one weekly series. The Federal Reserve Bank of Atlanta’s real-time nowcasting tool, GDPNow, uses a dynamic factor model combined with a GDP accounting approach to estimate current-quarter GDP.

There are other methods for nowcasting and mixed-frequency data. One popular tool for single-equation prediction using mixed-frequency data is the MIDAS model (Ghysels, Sinko, and Valkanov 2007), in which high-frequency data



are temporally aggregated using data-dependent weights. For a survey of methods of mixed-frequency nowcasting and forecasting, see Foroni and Marcellino (2013).

## Model Instability and Latent Variables

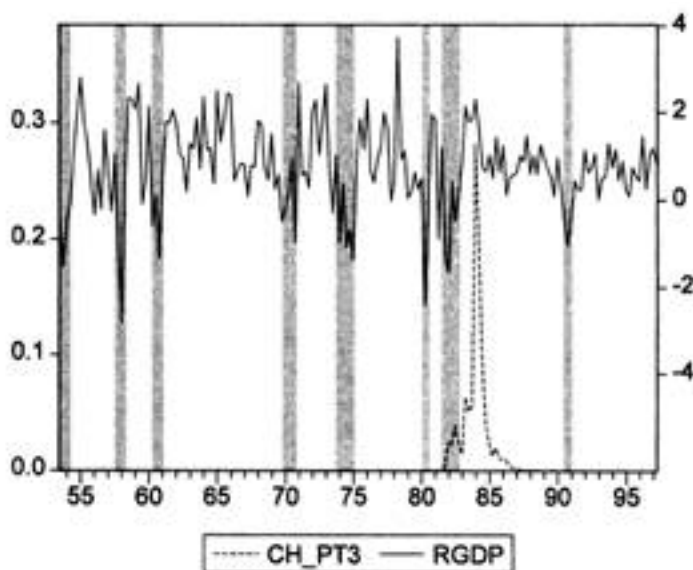
A large empirical literature has documented instability in both large- and small-dimensional time series models. A particularly well-known example of this instability is the Great Moderation, the period from 1984 to 2007 in which the volatility of many macroeconomic time series was greatly reduced. Examples of some of the many papers that document instability in the parameters of time series models include Stock and Watson (1996) for univariate time series forecasts, Stock and Watson (2003) for inflation forecasts using asset prices, and Welch and Goyal (2008) for equity premium forecasts. The methods in this literature draw in part on tests for breaks, time variation, and out-of-sample stability that date to the early 1990s.

This widespread nature of instability in time series relations raises the question of how to modify time series models so that they can be useful even in the presence of instability. An early approach was to model instability as deterministic regime shifts, but while useful, that approach is often unsatisfying because, outside of applications to a policy regime shift, the single-break model is an approximation and in any event there is rarely a reason to think that another shift will not occur. After all, the Great Moderation was followed by the financial crisis. A more appealing modeling strategy is to allow model parameters to evolve over time according to a stochastic process. If those time-varying parameters multiply observed variables, then the model has a linear state space (hidden Markov) structure and the Gaussian likelihood can be computed (using the Kalman filter). If, however, the time-varying parameters multiply latent variables, then it has an inherently nonlinear structure. Estimating such models is challenging, and it was clear 20 years ago that the rudimentary methods available needed to be improved.

The next two figures illustrate developments in the estimation of nonlinear latent variable models over the past 20 years. The first, Figure 6, is from Kim and Nelson (1999); the figure shows real GDP growth (the solid line), and the posterior probability of a break in the variance in GDP (dashed line). Based on this figure, Kim and Nelson (1999) concluded that US GDP growth had entered a period of low volatility, and that the most likely date for this transition was 1984Q1. This conclusion was reached independently using break test methods by McConnell and Perez-Quiros (2000). This low-volatility period, which lasted through 2007 (and to which the economy seems to have returned) subsequently became known as the Great Moderation.

Aside from its seminal empirical finding, Figure 6 illustrates a major methodological development in handling nonlinear and/or non-Gaussian time series models with latent variables. Kim and Nelson's (1999) model falls in this category: it allows for a one-time shift in the mean and variance of GDP growth, layered on top of Hamilton's (1989) stochastic regime shift model with recurrent shifts in the mean (which, Hamilton found, aligned with business cycle turning points). A challenge in these models is estimating the time path of the latent variable given all the data,

Figure 6

**US GDP Growth and the Posterior Probability of a Regime Change in its Innovation Variance**

Source: Kim and Nelson (1999), Fig. 3.A.

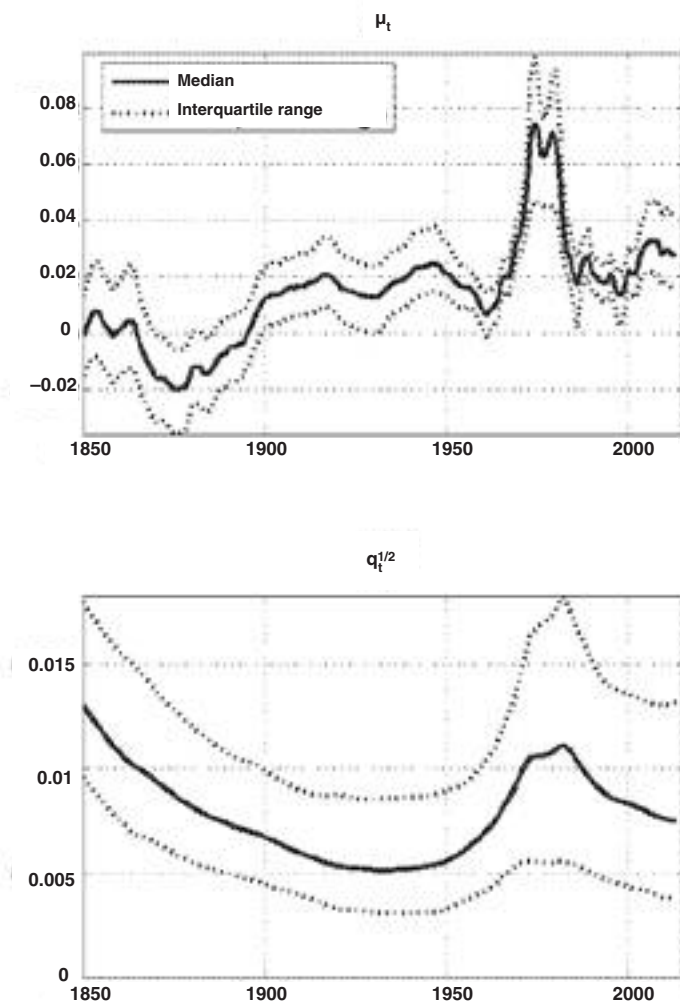
Note: The figure shows real GDP growth (the solid line), and the posterior probability of a break in the variance in GDP (dashed line). Based on this figure, Kim and Nelson (1999) concluded that US GDP growth had entered a period of low volatility, and that the most likely date for this transition was 1984Q1.

the so-called smoothing problem, along with the model parameters. To estimate their parameters and to solve this smoothing problem—to produce Figure 6—Kim and Nelson used Markov Chain Monte Carlo methods, which break down their complicated nonlinear non-Gaussian model into a sequence of Monte Carlo simulations using simpler models. Over the past 20 years, Markov Chain Monte Carlo has become a widely used tool for estimating seemingly intractable nonlinear/non-Gaussian models. With this tool, Kim and Nelson were able to obtain the posterior distribution of a one-time structural break in the variance which, as Figure 6 shows, strongly points to a reduction in the variance of GDP growth early in 1984.

The next figure, Figure 7, shows two panels from Cogley and Sargent (2015) that illustrate the incorporation of stochastic volatility into latent state variables. Cogley and Sargent use a univariate model that decomposes the rate of inflation into unobserved permanent and transitory (measurement error) components, both of which have innovations with time-varying variances. These variances are modeled as latent stochastic volatility processes. From a technical perspective, the situation is similar to that faced by Kim and Nelson (1999) in that the resulting model expresses the observed data as a nonlinear function of unobserved random variables (the permanent and transitory components of inflation and their volatilities). While the details differ, the Cogley–Sargent model is also readily estimated by Markov Chain Monte Carlo methods.

Figure 7

**Trend Inflation (Upper Panel) and the Standard Deviation of the Trend Innovation (Lower Panel) in an Unobserved Components–Stochastic Volatility Model of US Inflation, 1850–2012**



Source: Cogley-Sargent (2015), Fig. 7(A, C).

Note: Figure 7 illustrates the incorporation of stochastic volatility into latent state variables. Cogley and Sargent (2015) use an unobserved-components/stochastic-volatility model to study the evolution of the US inflation process from 1850 to 2012. Their posterior estimate of trend inflation is shown in the first panel, and their estimate of the time-varying standard deviation of changes in the trend is shown in the second panel. They find the periods of greatest variance in the trend to be during the Civil War and during the period of inflation and disinflation in the 1970s and early 1980s.

Cogley and Sargent (2015) use this unobserved-components/stochastic-volatility model to study the evolution of the US inflation process from 1850 to 2012. Their posterior estimate of trend inflation is shown in the first panel, and their estimate of the time-varying standard deviation of changes in the trend is shown in the second panel. They find the periods of greatest variance in the trend to be during the Civil War and during the period of inflation and disinflation in the 1970s and early 1980s.

The literature on nonlinear/non-Gaussian filtering is complex, nuanced, and massive. See Durbin and Koopman (2012) for a textbook treatment of linear and nonlinear filtering methods.

## More Reliable Inference

Finally, the past 20 years has seen important work that aims to improve the quality of statistical inferences. In the mid-1990s, several influential studies found that widely used methods for computing test statistics with time series data could reject far too often or, said differently, that confidence intervals could fail to include the true parameter value far less frequently than the claimed 95 percent coverage rate. Theoretical econometricians recognized that more work was needed, particularly in the areas of instrumental variables where the instrument might be weak, standard errors for regression with serially correlated errors, and regression with highly persistent regressors.

### Weak Instruments and Weak Identification

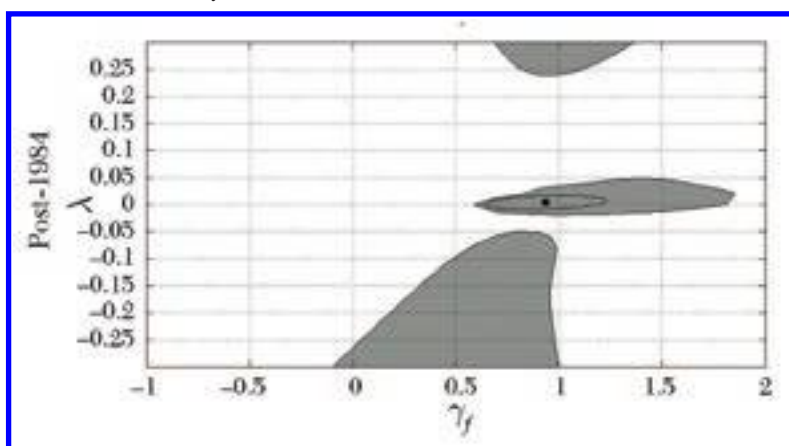
A weak instrument has a small correlation with the variable it is instrumenting, given the other included variables. For decades, conventional wisdom held that a weak instrument would simply produce large standard errors, which would correctly convey that the information in that variable is scant. But a series of papers in the 1990s showed that the consequences of a so-called weak instrument were more serious: the estimator will in general be biased, conventional standard errors are misleading, and these problems can occur in very large samples.<sup>3</sup> This problem, which is more generally referred to as weak identification, also arises in generalized method of moments estimation. Although weak instruments have received the most attention in microeconometrics, the inferential challenges posed by weak identification also have played a role in time series econometrics over the past 20 years.

The next figure, taken from Mavroeidis, Plagborg-Møller, and Stock (2014), illustrates the problems with using conventional asymptotic standard errors and confidence intervals in instrumental variables methods when one has weak instruments. Figure 8 shows confidence sets for two key parameters of the hybrid New Keynesian Phillips Curve; on the vertical axis,  $\lambda$  is the coefficient on marginal cost (or, in other specifications, the unemployment gap or output gap) and, on the

<sup>3</sup>Key papers on this subject from the 1990s include Nelson and Startz (1990a, 1990b) and Hansen, Heaton, and Yaron (1996) (Monte Carlo simulations), Bound, Jaeger, and Baker (1995) (empirical application), and Staiger and Stock (1997) (econometric theory).

Figure 8

**Point Estimate and 90% Confidence Sets for Hybrid New Keynesian Phillips Curve Parameters: Standard Generalized Method of Moments (Ellipse) and Weak-Instrument Robust (Gray)**



Source: Mavroeidis, Plagborg-Møller, and Stock (2014), Fig. 11a.

Note: Figure 8 shows confidence sets for two key parameters of the hybrid New Keynesian Phillips Curve. The dot is the point estimate using generalized method of moments, and the small ellipse around the point estimate is the corresponding nominal 90 percent confidence set computed using textbook asymptotics. The gray regions in the figure comprise a 90 percent confidence set that is robust to the use of weak instruments. The figures show that the weak-identification robust confidence sets differ dramatically from the standard asymptotic confidence ellipse. See text for details.

horizontal axis,  $\gamma_f$  is the coefficient on forward-looking rational expectations (sometimes interpreted as relating to the fraction of forward-looking agents). The results in this figure were computed using data from 1984–2011, where, following Galí and Gertler (1999), the labor share is the proxy for marginal cost, and the instruments are three lags each of marginal cost and the change in inflation, pruned down from Galí and Gertler's (1999) original set of 24 instruments (which yield similar qualitative results). The dot is the point estimate using generalized method of moments, and the small ellipse around the point estimate is the corresponding nominal 90 percent confidence set computed using textbook asymptotics. The gray regions in the figure comprise a 90 percent confidence set that is robust to the use of weak instruments. The obvious conclusion from Figure 8 is that the weak-identification robust confidence sets differ dramatically from the standard asymptotic confidence ellipse. Mavroeidis, Plagborg-Møller, and Stock (2014) argue that the reason for this divergence is that the instruments used in this generalized method of moments estimation are weak. This problem of weak identification arises broadly in New Keynesian Phillips Curve applications (for example, Henry and Pagan 2004; Mavroeidis 2004; Nason and Smith 2008).

Weak identification also arises in other contexts, like in the estimation of intertemporal consumption-based asset pricing models (Stock and Wright 2000) and estimation of monetary policy reaction functions using generalized method of moments (Consolo and Favero 2009). Weak identification arises in some types of

inference in structural autoregressions (for example, Pagan and Robertson 1998; Chevillon, Mavroeidis, and Zhan 2016; for more references, see Stock and Watson 2016, Section 4). It also arises in complicated ways in the estimation of dynamic stochastic equilibrium models (for example, Andrews and Mikusheva 2015; Qu 2014).

In linear instrumental variable regressions, one commonly used diagnostic is to check if the  $F$ -statistic testing the hypothesis that the coefficient(s) on the instrument(s) in the first stage of two stage least squares—the so-called first-stage  $F$ -statistic—is less than 10; if so, weak identification is potentially a problem. This specific approach is specialized to the homoskedastic setting with uncorrelated errors; approaches to extending this to heteroskedasticity are proposed by Montiel Olea and Pflueger (2013) and Andrews (2016).

In the simplest models—the textbook regression model with a single endogenous regressor and errors that are homoskedastic and serially uncorrelated—there are now methods for dealing with weak instruments with very good size and power, both asymptotically and in finite samples. As one departs from this model, most notably when the number of parameters gets large and/or the model is nonlinear in the parameters, the toolkit is less complete and theoretical work remains under way.

### **Inference with Serially Correlated and Potentially Heteroskedastic Errors**

In time series data with a serially correlated error term, each additional observation does not provide entirely new information about the regression coefficient. Moreover, many time series regressions exhibit clear signs of heteroskedasticity. In this setting, the ordinary least squares standard error formula does not apply and instead standard errors that are robust to heteroskedasticity and autocorrelation must be used. For example, this problem arises when the dependent variable is a multi-period return or a multiple-period-ahead variable. The problems of heteroskedasticity and autocorrelation also arise in generalized method of moments models when the data are serially correlated.

In practice, the most commonly used standard errors that are heteroskedasticity- and autocorrelation-robust are computed using methods from seminal papers by Newey and West (1987) and Andrews (1991). These methods compute standard errors by replacing the estimate of the variance of the product of the regressor and the error in the usual heteroskedasticity-robust formula for the variance of the ordinary least squares estimator with a weighted average of the autocovariances of that product; the number of autocovariances averaged is determined by the so-called “bandwidth” parameter. But even 20 years ago, there were inklings that the performance of hypothesis tests and confidence intervals constructed using these standard errors in typical macroeconomic applications fell short of the asymptotic performance used to justify the tests. In an early Monte Carlo simulation, den Haan and Levin (1997) studied the rejection rates of tests using these standard errors under the null hypothesis—that is, the size of the test. Depending on the persistence in the data, they found that a test that should reject 5 percent of the time under the null will in practice reject 10 or even 20 percent of the time. If the aim of a research project is, say, to test for predictability in multiyear stock returns using monthly

data, this over-rejection could easily lead to an incorrect conclusion that returns are predictable when in fact they are not.

Understanding the source of these size distortions and improving upon Newey–West/Andrews standard errors therefore became a major line of research by theoretical econometricians over the past 20 years, which is succinctly surveyed by Müller (2014, Sections 2–3). In brief, this line of work finds that to construct tests with a rejection rate closer to the desired 5 percent, it is necessary to use bandwidths much larger than those suggested by Newey–West and Andrews. But doing so results in a complication: the test statistic no longer has the usual large-sample normal distribution and, in general, nonstandard critical values must be used. These ideas were set out by Kiefer, Vogelsang, and Bunzel (2000), and their insights prompted a large literature aimed at understanding and refining their large-bandwidth approach. This theoretical literature has now produced multiple methods that yield far smaller size distortions than tests based on Newey–West/Andrews standard errors, and which also have better power than the Kiefer–Vogelsang–Bunzel test. Moreover, some of these tests have standard critical values, simplifying their use in practice.

Applied econometricians typically are eager to use the most recent econometric method when they demonstrably improve upon the methods of the past. Curiously, this has not been the case for heteroskedasticity- and autocorrelation-robust inference, where empirical practice continues to be dominated by Newey–West/Andrews standard errors. The new methods are easy to use, straightforward to understand, and have a lineage that traces back 40 years. It is time for empirical researchers in time series econometrics to take the next step and to adopt these improved methods for heteroskedasticity- and autocorrelation-robust inference.

### **Long-run Relations, Cointegration, and Persistent Regressors**

The basic insight of cointegration—the development for which Clive Granger received the Nobel Prize in 2003—is that multiple persistent macroeconomic variables move together at low frequencies, that is, they share common long-term trends. Moreover, these low-frequency comovements connect with basic economic theories such as balanced economic growth. But while there was a surge of work on cointegration in the 1980s and 1990s, such work has received less emphasis since then.

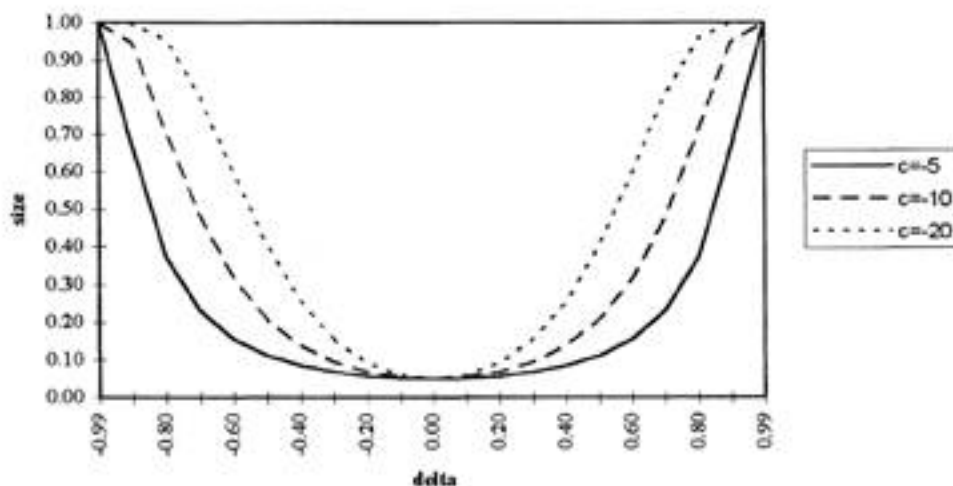
Our final historical figure, from Elliott (1998), illustrates a technical roadblock hit by this research program. Elliott’s figure, our Figure 9, portrays the null rejection rate of a test of the value of a cointegrating coefficient in a simple model with two cointegrated variables. The test maintains that each of the variables is integrated of order one, that is, has a unit autoregressive root, an assumption that is part of the cointegration model. Figure 9 shows that small departures from this unit-root assumption (as measured by  $c$ , which is the difference between the true largest root and one, multiplied by the sample size) can cause major problems for tests and confidence intervals about the value of that cointegrating coefficient: tests that are supposed to reject 5 percent of the time under the null can reject with very high rates (shown on the vertical axis), particularly when the correlation  $\delta$  (shown on the horizontal axis) between innovations in the error and in



Figure 9

**Asymptotic Size of Tests of Values of the Cointegrating Coefficient Using Efficient Cointegrating Estimators and Their Standard Errors when the Time Series Follow Local-to-Unity Processes with Parameter  $c$**

( $\delta$  is the correlation between innovations in the error and in the regressor)



Source: Elliott (1998), Figure 1(a).

*Note:* The figure portrays the null rejection rate of a test of the value of a cointegrating coefficient in a simple model with two cointegrated variables. The test maintains that each of the variables is integrated of order one, that is, has a unit autoregressive root, an assumption that is part of the cointegration model. The figure shows that small departures from this unit-root assumption (as measured by  $c$ , which is the difference between the true largest root and one, multiplied by the sample size) can cause major problems for tests and confidence intervals about the value of that cointegrating coefficient.

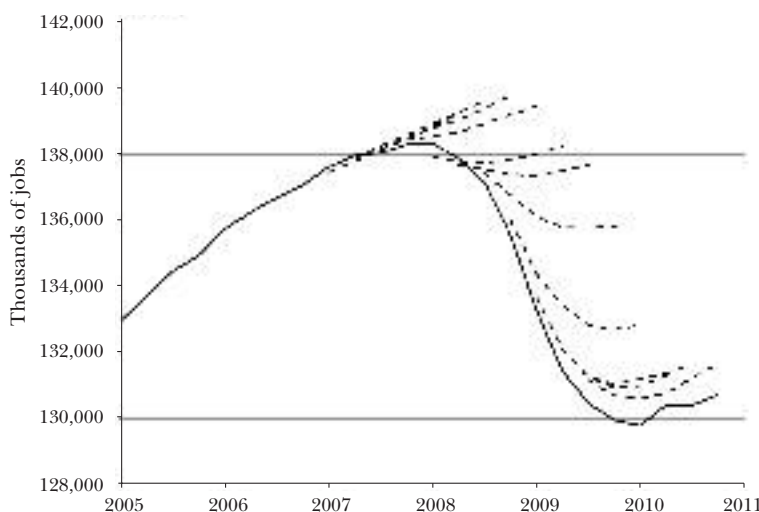
the regressor is large. In fact, this problem arises for deviations from a unit root that are too small to be detected with high probability, even in arbitrarily large samples. As a result, standard methods of inference developed for cointegration models are not robust to effectively undetectable departures from the model, making such inference unreliable.

While subsequent work has produced novel ideas by econometric theorists, the proposed methods have drawbacks and no alternative set of procedures have emerged. In fact, the literature has shown that the problem documented in Figure 9 goes beyond the local-to-unity model used by Elliott (1998) and other researchers in this area. Related problems of inference also arise in regressions in which a regressor is persistent, as can occur in applications with financial data.

It is important to stress that these challenges are technical ones; the basic insight of cointegration that variables move together at low frequencies is a deep one that connects with core economic theories such as balanced growth and the term structure of interest rates. But inference, and perhaps modeling, of those comovements can be more complicated than had originally been thought.

*Figure 10*

**“The Mother of All Forecast Errors”: Survey of Professional Forecasters Median Forecast for Nonfarm Business Employment during the 2007–2009 Recession and Early Recovery**



Source: Philadelphia Fed Survey of Professional Forecasters.

Note: This figure shows the real-time median forecast of the log of nonfarm employment recorded by the Survey of Professional Forecasters in the quarters leading up to and through the financial crisis. Even well after the crisis began and real-time information about the collapse of the economy was available, these forecasters consistently predicted a mild recession.

## Challenges Ahead

We close by mentioning a few of the research challenges for time series econometrics. Our final figure shows that despite the substantial improvements in forecasting methods over the past decades, much work remains. When we teach, we call Figure 10 the “Mother of All Forecast Errors.” This figure shows the real-time median forecast of the log of nonfarm employment recorded by the Survey of Professional Forecasters in the quarters leading up to and through the financial crisis. Even well after the crisis began and real-time information about the collapse of the economy was available, these forecasters consistently predicted a mild recession. A small part of these errors is due to revisions between preliminary and final data, but most of these errors, we believe, represent a failure of forecasting models to capture the severity of the shocks and their devastating effect on the economy. Forecasters certainly were not the only economists to misjudge events leading up to and during the financial crisis! But this is an article about time series methods, and in our view, tackling the challenge of Figure 10 is a priority.

Another open challenge lies in the big data sphere. The methods of the past 20 years—dynamic factor models and large Bayesian vector autoregressions—have made it possible to include arbitrarily many series in forecasting systems and to incorporate data releases in real time, and the result has been large improvements in

macroeconomic monitoring. However, there is some evidence that the parametric restrictions (or priors) that make these methods work discard potentially important information. In the context of dynamic factor models, the question is whether there is useful information in the higher factors beyond the handful that would normally be included (such as the six factors used to produce Figure 3). Some studies have looked at this question, with mixed results; for example, Carrasco and Rossi (2016) give some positive results, while we give some negative results in Stock and Watson (2012b). A more ambitious question is whether there is exploitable nonlinear structure in these data that could perhaps be revealed by modern machine learning methods. While it is tempting to dive in and use a battery of machine learning methods to attack these data, one must remember that data snooping can lead to unintentional overstatement of results. One advantage of dynamic factor models, after all, is that they are closely linked to dynamic macro models (Sargent 1989; Boivin and Giannoni 2006). We suspect that the next steps towards exploiting additional information in large datasets will need to use new statistical methods guided by economic theory.

Separately, there are important open questions relating to low-frequency time series econometrics. For example, what does historical evidence tell us about whether the recent slowdown in US productivity is permanent or temporary? The answer to this question is crucial for many long-term economic issues, such as the future of Social Security and valuing policies to mitigate climate change. Another, technically related set of questions returns to the basic insight of cointegration and the challenge posed by Elliott's (1989) figure (Figure 9): there are clearly low-frequency comovements in the data, and macroeconometricians need a set of tools for quantifying those comovements that does not hinge on adopting a particular model, such as a unit root model, for the underlying trends. These are technically difficult problems, and Müller and Watson (2016a, 2016b) propose possible avenues for tackling them.

Finally, there are a number of opportunities for expanding identification and estimation of macro models by using information in microeconomic data. Here, opportunities range from estimation of parameters describing individual preferences and firm behavior, to the possibility of using rich micro data to improve macro monitoring and forecasting.

The earliest empirical work in macroeconomics relied on time series data; indeed the first instrumental variables regression was estimated in 1926 using time series data. The past 20 years has seen a continuation of the vigorous development of methods for using time series data. These methods draw on improved computational capacity, better data availability, and new understandings in econometric and statistical theory. The core driver of these developments is the need of policymakers for reliable guidance on the effects of contemplated policies, along with their shared need with the private sector to understand where the economy is and where it is going. Those needs will not go away. If anything, they become more urgent in our volatile and ever-changing economic environment. Although the challenges facing time series econometricians are difficult, so have they been in the past, and exciting and highly relevant research programs beckon.

■ *We thank Gray Calhoun, Mikkel Plagborg-Møller, and the authors of the papers from which we took figures, for helpful comments.*

## References

- Ait-Sahalia, Yacine, and Lars Peter Hansen. 2010. *Handbook of Financial Economics*, Vol. 1: *Tools and Techniques* and Vol. 2: *Applications*. Elsevier.
- Altissimo, Filippo, Riccardo Cristadoro, Mario Forni, Marco Lippi, and Giovanni Veronese. 2010. "New EuroCOIN: Tracking Economic Growth in Real Time." *Review of Economics and Statistics* 92(4): 1024–34.
- Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59(3): 817–858.
- Andrews, Isaiah. 2016. "Valid Two-Step Identification-Robust Confidence Sets for GMM." <http://economics.mit.edu/files/11848>.
- Andrews, Isaiah, and Anna Mikusheva. 2015. "Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models." *Quantitative Economics* 6(1): 123–52.
- Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti. 2009. "Real-Time Measurement of Business Conditions." *Journal of Business & Economic Statistics* 27(4): 417–27.
- Bai, Jushan. 2003. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71(1): 135–72.
- Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70(1): 191–21.
- Bai, Jushan, and Serena Ng. 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica* 74(4): 1133–50.
- Bank of Norway. 2016. "Monetary Policy Report with Financial Stability Assessment 4/16." December 15, 2016. <http://www.norges-bank.no/en/Published/Publications/Monetary-Policy-Report-with-financial-stability-assessment/416-Monetary-Policy-Report/>.
- Bañbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin. 2013. "Now-Casting and the Real-Time Data Flow." Chap. 4 in *Handbook of Economic Forecasting*, vol. 2, edited by Graham Elliott, and Alan Timmermann. Elsevier, North-Holland.
- Baumeister, Christiane, and James D. Hamilton. 2015. "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information." *Econometrica* 83(5): 1963–99.
- Bernanke, Ben S., Jean Boivin, and Piotr Elias. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120(1): 387–422.
- Bernanke, Ben S., and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60(3): 1221–57.
- Boivin, Jean, and Marc Giannoni. 2006. "DSGE Models in a Data-Rich Environment." NBER Working Paper 12772.
- Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90: 443–50.
- Burns, Arthur F., and Wesley C. Mitchell. 1946. *Measuring Business Cycles*. NBER.
- Canova, Fabio. 2007. *Methods for Applied Macroeconomic Research*. Princeton University Press.
- Chevillon, Guillaume, Sophocles Mavroeidis, and Zhaoguo Zhan. 2016. "Robust Inference in Structural VARs with Long-Run Restrictions." <https://sites.google.com/site/sophoclesmavroeidis/research/lrsvar.pdf?attredirects=0>.
- Chudick, Alexander, and M. Hashem Pesaran. 2016. "Theory and Practice of GVAR Modelling." *Journal of Economic Surveys* 30(1): 165–97.
- Cochrane, John H., and Monica Piazzesi. 2002. "The Fed and Interest Rates: A High-Frequency Identification." *American Economic Review* 92(2): 90–95.
- Cogley, Timothy, and Thomas J. Sargent. 2015. "Measuring Price-Level Uncertainty and Instability in the United States, 1850–2012." *Review of Economics and Statistics* 97(4): 827–38.
- Consolo, Agostino, and Carlo A. Favero. 2009. "Monetary Policy Inertia: More a Fiction than a Fact?" *Journal of Monetary Economics* 56(6): 900–906.
- Corradi, Valentina, and Norman R. Swanson. 2006. "Predictive Density Evaluation." Chap. 5 in *Handbook of Economic Forecasting*, vol. 1, edited by Graham Elliott, Clive W. J. Granger, and Allan Timmermann. Elsevier: North Holland.
- Corrasco, Marine, and Barbara Rossi. 2016. "In-Sample Inference and Forecasting in Misspecified Factor Models." *Journal of Business and Economic Statistics* 34(3): 313–38.
- DeJong, David N., Beth F. Ingram, and Charles H. Whiteman. 2000. "A Bayesian Approach to Dynamic Macroeconomics." *Journal of Econometrics* 98(2): 203–223.
- den Haan, W. J., and A. Levin. 1997. "A Practitioners Guide to Robust Covariance Matrix Estimation." Chap. 12 of *Handbook of Statistics*, Vol. 15: *Robust Inference*, edited by G. S. Maddala and C. R. Rao. Elsevier.
- Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*, 2nd edition. Oxford University Press.
- Elliott, Graham. 1998. "On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots." *Econometrica* 66(1): 149–58.
- Elliott, Graham, and Alan Timmermann, eds. 2013. *Handbook of Economic Forecasting*, vol. 2. Elsevier.
- Elliott, Graham, and Alan Timmermann. 2016. *Economic Forecasting*. Princeton University Press.

- Faust, Jon, John H. Rogers, Eric Swanson, and Jonathan H. Wright. 2003. "Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data." *Journal of the European Economic Association* 1(5): 1031–57.
- Federal Reserve Bank of New York. 2017. "Nowcasting Report, February 10, 2017." Available at: <https://www.newyorkfed.org/research/policy/nowcast>.
- Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, and Frank Schorfheide. 2016. "Solution and Estimation Methods for DSGE Models." Chap. 9 in *Handbook of Macroeconomics*, vol. 2, edited by John B. Taylor and Harald Uhlig. Elsevier.
- Forni, Claudia, and Massimiliano Marcellino. 2013. "A Survey of Econometric Methods for Mixed-Frequency Data." Norges Bank Research Working Paper 2013-6.
- Friedman, Walter F. 2009. "The Harvard Economic Service and the Problems of Forecasting." *History of Political Economy* 41(1): 57–88.
- Fry, Renée, and Adrian Pagan. 2011. "Sign Restrictions in Structural Vector Autoregressions: A Critical Review." *Journal of Economic Literature* 49(4): 938–60.
- Gali, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44(2): 195–222.
- Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7(1): 44–76.
- Ghysels, Eric, Arthur Sinko, and Rossen Valkanov. 2007. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews* 26(1): 53–90.
- Giacomini, Raffaella, and Toru Kitagawa. 2014. "Inference about Non-Identified SVARs." CEPR Discussion Paper 10287.
- Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57(2): 357–84.
- Hamilton, James D. 2003. "What Is an Oil Shock?" *Journal of Econometrics* 113(2): 363–98.
- Hansen, Lars P., John Heaton, and Amir Yaron. 1996. "Finite Sample Properties of Some Alternative GMM Estimators." *Journal of Business and Economic Statistics* 14(3): 262–80.
- Henry, S. G. B., and A. R. Pagan. 2004. "The Econometrics of the New Keynesian Policy Model: Introduction." *Oxford Bulletin of Economics and Statistics* 66(Supplement): 581–607.
- Herbst, Edward P., and Frank Schorfheide. 2015. *Bayesian Estimation of DSGE Models*. Princeton University Press.
- Ireland, Peter N. 1997. "A Small, Structural, Quarterly Model for Monetary Policy Evaluation." *Carnegie-Rochester Conference Series on Public Policy* 47: 83–108.
- Kiefer, Nicholas M., Timothy J. Vogelsang, and Helle Bunzel. 2000. "Simple Robust Testing of Regression Hypotheses." *Econometrica* 68(3): 695–714.
- Kilian, Lutz. 2008. "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics* 90(2): 216–40.
- Kim, Chang-jin, and Charles R. Nelson. 1999. "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle." *Review of Economics and Statistics* 81(4): 608–616.
- Kuttner, Kenneth N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics* 47(3): 523–44.
- Lütkepohl, Helmut. 2013. "Identifying Structural Vector Autoregressions via Changes in Volatility." *Advances in Econometrics*, vol. 32, pp. 169–203.
- Mavroeidis, Sophocles. 2004. "Weak Identification of Forward-Looking Models in Monetary Economics." *Oxford Bulletin of Economics and Statistics* 66(Supplement): 609–35.
- Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52(1): 124–88.
- McConnell, Margaret M., and Gabriel Perez-Quiro. 2000. "Output Fluctuations in the United States: What Has Changed Since the Early 1980's." *American Economic Review* 90(5): 1464–76.
- McCracken, Michael W., and Serena Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics* 34(4): 574–89.
- Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103(4): 1212–47.
- Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31(3): 358–69.
- Moon, Hyungsik Roger, Frank Schorfheide, and Eleonara Granziera. 2013. "Inference for VARs Identified with Sign Restrictions." <http://sites.sas.upenn.edu/schorf/files/svar-paper.pdf>.
- Müller, Ulrich K. 2014. "HAC Corrections for Strongly Autocorrelated Time Series." *Journal of Business and Economic Statistics* 32(3): 311–22.
- Müller, Ulrich K., and Mark W. Watson. 2016a. "Measuring Uncertainty about Long-Run Predictions." *Review Economic Studies* 84(4): 1711–40.
- Müller, Ulrich K., and Mark W. Watson. 2016b. "Long-Run Covariability." Unpublished paper, Princeton University.
- Nason, James M., and Gregor W. Smith. 2008. "Identifying the New Keynesian Phillips Curve." *Journal of Applied Econometrics* 23(5): 525–51.



- Nelson, Charles R., and Richard Startz.** 1990a. "The Distribution of the Instrumental Variable Estimator and Its  $t$  Ratio When the Instrument Is a Poor One." *Journal of Business* 63(1, Part 2): S125–S140.
- Nelson, Charles R., and Richard Startz.** 1990b. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica* 58(4): 967–76.
- Newey, Whitney K., and Kenneth D. West.** 1987. "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–708.
- Normandin, Michel, and Louis Phaneuf.** 2004. "Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility." *Journal of Monetary Economics* 51(6): 1217–43.
- Otrok, Christopher.** 2001. "On Measuring the Welfare Costs of Business Cycles." *Journal of Monetary Economics* 47(1): 61–92.
- Pagan, A. R., and J. C. Robertson.** 1998. "Structural Models of the Liquidity Effect." *Review of Economics and Statistics* 80(2): 202–217.
- Plagborg-Møller, Mikkel.** 2016. "Bayesian Inference on Structural Impulse Response Functions." [http://scholar.harvard.edu/files/plagborg/files/irf\\_bayes.pdf](http://scholar.harvard.edu/files/plagborg/files/irf_bayes.pdf).
- Qu, Zhongjun.** 2014. "Inference in DSGE Models with Possible Weak Identification." *Quantitative Economics* 5(2): 457–94.
- Ramey, V. A.** 2016. "Macroeconomic Shocks and their Propagation." Chap. 2 in *Handbook of Macroeconomics* vol. 2., edited by J. B. Taylor and H. Uhlig, 71–162. Elsevier.
- Rigobon, Roberto.** 2003. "Identification through Heteroskedasticity." *Review of Economics and Statistics* 85(4): 777–92.
- Rigobon, Roberto, and Brian Sack.** 2003. "Measuring the Reaction of Monetary Policy to the Stock Market." *Quarterly Journal of Economics* 118(2): 639–69.
- Rigobon, Roberto, and Brian Sack.** 2004. "The Impact of Monetary Policy on Asset Prices." *Journal of Monetary Economics* 51(8): 1553–75.
- Romer, Christina D., and David H. Romer.** 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." In *NBER Macroeconomics Annual 1989*, vol. 4, edited by Olivier Blanchard and Stanley Fisher, 121–70. MIT Press.
- Rudebusch, Glenn D.** 1998. "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review* 39(4): 907–931.
- Sargent, Thomas J.** 1989. "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy* 97(2): 251–87.
- Schorfheide, Frank.** 2000. "Loss Function-based Evaluation of DSGE Models." *Journal of Applied Econometrics* 15(6): 645–70.
- Sentana, Enrique, and Gabriele Fiorentini.** 2001. "Identification, Estimation, and Testing of Conditionally Heteroskedastic Factor Models." *Journal of Econometrics* 102(2): 143–64.
- Sims, Christopher A.** 1980. "Macroeconomics and Reality." *Econometrica* 48(1): 1–48.
- Smets, Frank, and Raf Wouters.** 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1(5): 1123–75.
- Staiger, Douglas, and James H. Stock.** 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3): 557–86.
- Stock, James H.** 2008. *What's New in Econometrics: Time Series, Lecture 7*. Short course lectures, NBER Summer Institute, at [http://www.nber.org/minicourse\\_2008.html](http://www.nber.org/minicourse_2008.html).
- Stock, James H., and Mark W. Watson.** 1989. "New Indexes of Coincident and Leading Economic Indicators." *NBER Macroeconomics Annual 1989*, edited by Olivier J. Blanchard and Stanley Fischer, 351–93. MIT Press.
- Stock, James H., and Mark W. Watson.** 1996. "Evidence on Structural Instability in Macroeconomic Time Series Relations." *Journal of Business and Economic Statistics* 14(1): 11–30.
- Stock, James H., and Mark W. Watson.** 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97(460): 1167–79.
- Stock, James H., and Mark W. Watson.** 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* 41(3): 788–829.
- Stock, James H., and Mark W. Watson.** 2011. *Introduction to Econometrics*, 3rd Edition. Pearson.
- Stock, James H., and Mark W. Watson.** 2012a. "Disentangling the Channels of the 2007–09 Recession." *Brookings Papers on Economic Activity*, no. 1, 81–135.
- Stock, James H., and Mark W. Watson.** 2012b. "Generalized Shrinkage Methods for Forecasting Using Many Predictors." *Journal of Business & Economic Statistics* 30(4): 481–93.
- Stock, James H., and Mark W. Watson.** 2016. "Dynamic Factor Models, Factor-Augmented Autoregressions, and Structural Vector Autoregressions in Macroeconomics." Chap. 8 in *Handbook of Macroeconomics*, vol. 2, edited by John B. Taylor and Harald Uhlig, 415–526. Elsevier.
- Stock, James H., and Jonathan H. Wright.** 2000. "GMM with Weak Identification." *Econometrica* 68(5): 1055–96.
- Uhlig, Harald.** 2005. "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics* 52(2): 381–419.
- Welch, Ivo, and Amit Goyal.** 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies* 21(4): 1455–1508.