JAMES H. STOCK
Harvard University

MARK W. WATSON
Princeton University

Disentangling the Channels of the 2007–09 Recession

ABSTRACT   This paper examines the macroeconomic dynamics of the 2007–09 recession in the United States and the subsequent slow recovery. Using a dynamic factor model with 200 variables, we reach three main conclusions. First, although many of the events of the 2007–09 collapse were unprecedented, their net effect was to produce macro shocks that were larger versions of shocks previously experienced, to which the economy responded in a historically predictable way. Second, the shocks that produced the recession were primarily associated with financial disruptions and heightened uncertainty, although oil shocks played a role in the initial slowdown, and subsequent drag was added by effectively tight conventional monetary policy arising from the zero lower bound. Third, although the slow nature of the recovery is partly due to the shocks of this recession, most of the slow recovery in employment, and nearly all of the slow recovery in output, is due to a secular slowdown in trend labor force growth.

The recession that began in the fourth quarter of 2007 was unprecedented in the postwar United States for its severity and duration. Following the cyclical peak of 2007Q4, real GDP dropped by 5.1 percent and nearly 8.8 million jobs were lost. According to the most recent data revisions, that peak in GDP was not reattained until 15 quarters later, in 2011Q3, and as of this writing in April 2012 only 3.7 million jobs have been regained. All this suggests that the 2007–09 recession and subsequent recovery were qualitatively, as well as quantitatively, different from previous postwar recessions. The recession also seems unprecedented in its precipitating sources: the first persistent nationwide decline in real estate values since World War II, a financial sector that was unusually vulnerable because of recent deregulation

and little-understood derivatives, and a collapse in lending that also dampened the recovery.[1]

This paper takes an empirical look at this recession and recovery, with an eye toward quantifying the extent to which this recession differs from previous postwar recessions, the contributions of various shocks to the recession, and the reasons for the slow recovery. More specifically, we consider three questions. First, beyond its severity, how did this recession differ from previous postwar recessions? Second, what were the economic shocks that triggered this recession, and what were their quantitative contributions to the collapse of economic activity? Third, to what extent does the current "jobless" recovery constitute a puzzle, something out of line with historical patterns and thus requiring a new explanation?[2]

The organizing framework for our analysis of these three questions is a high-dimensional dynamic factor model (DFM). Like a vector autoregression (VAR), a DFM is a linear time-series model in which economic shocks drive the comovements of the variables; the main difference between a DFM and a VAR is that the number of shocks does not increase with the number of series. Also as in a VAR, some properties, such as stability and forecasts, can be studied using a "reduced form" DFM that does not require identifying factors or structural shocks; however, attributing movements in economic variables to specific economic shocks requires identifying those shocks as in structural VAR analysis.

Our empirical model has 200 macroeconomic variables driven by six macro factors. These six factors drive the macro comovements of all the variables. Shocks that affect only a handful of series, such as a sectoral demand shock that affects a small number of employment and production series, would not surface as a macro factor but would instead imply idiosyncratic variation in those series. Using this model, we can address the question of whether the 2007–09 recession was characterized by a new type of shock by examining whether that recession is associated with new factors.

Our three main findings follow the three questions posed above. First, a combination of visual inspection and formal tests of a DFM estimated

1. The view that financial recessions and recoveries are different from "normal" recessions has been articulated most notably by Reinhart and Rogoff (2009); see also Reinhart and Reinhart (2010), Hall (2010), Mishkin (2010), Bank of Canada (2011), and Jordà, Schularick, and Taylor (2011).

2. Various reasons have been proposed for why this recovery is exceptional, including the deleveraging that followed the financial crisis (for example, Mian and Sufi 2011), regional or industry job mismatch (for example, Şahin and others 2011), changes in labor management practices (for example, Berger 2011), and monetary policy rendered ineffective because of the zero lower bound.

through 2007Q3 suggests that the same six factors that explained previous postwar recessions also explain the 2007–09 recession: no new "financial crisis" factor is needed. Moreover, the response of the macro variables to these "old" factors is, in most cases, the same as it was in earlier recessions. Within the context of our model, the recession was associated with exceptionally large movements in these "old" factors, to which the economy responded predictably given historical experience. Of course, this recession *did* have new and unprecedented failures of the financial plumbing (for example, the fall of the global investment bank Lehman Brothers) as well as novel, aggressive policy responses (such as the Troubled Asset Relief Program, or TARP, and the Federal Reserve's various facilities); our results suggest, however, that these failures and responses did not have qualitatively different net effects on the macroeconomy than did past disturbances— just larger ones. What the results here suggest is that the shocks arising from these extraordinary events had ordinary impacts: from a macro perspective these shocks had the same effect as previously observed shocks, so the shocks surface in our model as large movements in the "old" factors.

Second, identifying what, precisely, these large economic shocks were requires an exercise similar to structural VAR identification. We consider six shocks: to oil markets, monetary policy, productivity, uncertainty, liquidity and financial risk, and fiscal policy. We identify these shocks using a novel method in which we treat shocks constructed elsewhere in the literature as instrumental variables; for example, one of the three instruments we use to identify the oil shock is Lutz Kilian's (2008a) series on strife-driven OPEC oil production shortfalls. In all, we have 17 such external instruments with which to estimate our six shocks. The results of this exercise are mixed, in large part because the instruments produce estimates of purportedly different shocks that are correlated. In particular, the uncertainty shocks and the shocks to liquidity and financial risk are highly correlated, which makes their separate interpretation problematic. Despite these drawbacks, the structural analysis is consistent with the recession being caused by initial oil price shocks followed by large financial and uncertainty shocks.

Third, focusing on the recovery following the 2009Q2 trough, we estimate that slightly less than half of the difference between the recovery in employment since 2009Q2 and the average for recoveries between 1960 and 1982 is attributable to cyclical factors (the shocks, or factors, during the recession). Instead, most of the slowness of the recovery is attributable to a long-term slowdown in trend employment growth. Indeed, that slowdown has been dramatic: according to our estimates, trend annual employment growth slowed from 2.4 percent in 1965 to 0.9 percent in 2005. The

explanation for this declining trend growth rate that we find most compelling rests on changes in underlying demographic factors, primarily the plateau in the female labor force participation rate (after a sharp rise during the 1970s through the 1990s) and the aging of the workforce. Because the net change in trend productivity growth over this period is small, this slower trend growth in employment translates into slower trend GDP growth. These demographic changes imply continued low or even declining trend growth rates in employment, which in turn suggest that future recessions will be deeper and longer, and will have slower recoveries, than has been the case historically.

A vast number of papers examine the financial crisis, but relatively few tackle the empirical macro issues discussed here. Some related papers that look at aspects of the problem of shocks and their propagation include Martin Lettau and Sydney Ludvigson (2011) on permanent wealth shocks; the related paper by John Campbell, Stefano Giglio, and Christopher Polk (2010) on reasons for the stock market collapse; Simon Gilchrist, V. Yankov, and Egon Zakrajšek (2011) on credit spreads and their role as measures of financial distress in this and previous recessions; and Robert Hall (2011, 2012) on the postcrisis dynamics. Òscar Jordà, Moritz Schularick, and Alan Taylor (2011) and Michael Bordo and Joseph Haubrich (2011) look at the relationship between the depth and the duration of recessions with a focus on whether financial crises are exceptional and reach opposite conclusions. We are not aware of a comprehensive treatment along the lines discussed here, however.

Section I of the paper describes the DFM and the data set. Section II presents a counterfactual exercise, examining how well the historical shocks and model do at predicting the 2007Q4–2011 experience, along with stability tests. Section III discusses identification of the structural shocks and provides empirical analysis of the identified shocks. Section IV focuses on the slow recovery, and section V concludes. Detailed data description and additional empirical results are contained in the online appendix.[3]

I. Empirical Methods and Data

We begin by describing our empirical methods and the data used in the analysis.

---

3. Online appendixes and replication files for the papers in this volume may be accessed on the Brookings Papers website, www.brookings.edu/about/projects/bpea, under "Past Editions." They are also accessible at Mark Watson's personal website at Princeton University, www.princeton.edu/~mwatson/.

I.A. Empirical Methods

Dynamic factor models capture the notion that the macroeconomy is driven by a handful of unobserved macro shocks. There is considerable empirical evidence that a DFM with a small number of factors describes the comovements of macroeconomic time series (see, for example, Sargent and Sims 1977, Giannone, Reichlin, and Sala 2004). Thomas Sargent (1989) and Jean Boivin and Marc Giannoni (2010) develop this idea formally, starting from a dynamic stochastic general equilibrium model in which the driving variables are observed with measurement error. There is now a rich set of econometric methods for inference in DFMs (see Stock and Watson 2011 for a survey). Applications of these methods include forecasting (see Eickmeier and Ziegler 2008) and the factor-augmented vector autoregression (FAVAR) method of Ben Bernanke, Boivin, and Piotr Eliasz (2005).

Because the comovements of the observed series stem from the factors, it is not necessary to model directly the dynamics among observed variables; this avoids the proliferation of coefficients found in VARs. Because a DFM has relatively few factors compared with the number of observed variables, it allows a tractable simultaneous empirical analysis of very many variables in a single, internally consistent framework.

THE DYNAMIC FACTOR MODEL Let $X_t = (X_{1t}, \ldots, X_{nt})'$ denote a vector of n macroeconomic time series observed over periods $t = 1, \ldots, T$, where X_{it} is a single time series, where all series have been transformed to be stationary and to have a mean of zero (details below), and let F_t denote the vector of r unobserved factors. The DFM expresses each of the n time series as a component driven by the factors, plus an idiosyncratic disturbance term e_{it}:

$$(1) \qquad X_t = \Lambda F_t + e_t,$$

where $e_t = (e_{1t}, \ldots, e_{nt})'$ and Λ is an $n \times r$ matrix of coefficients called the factor loadings. The term ΛF_t is called the "common component" of X_t.

The factors are modeled as evolving according to a VAR:

$$(2) \qquad \Phi(L)F_t = \eta_t,$$

where $\Phi(L)$ is an $r \times r$ matrix of lag polynomials and η_t is an $r \times 1$ vector of innovations.[4] Because the factor VAR in equation 2 is assumed to be stationary, F_t has the moving-average representation $F_t = \Phi(L)^{-1}\eta_t$. The idiosyncratic errors e_t can be serially correlated, but the methods used here do not require a parametric model of the e_t dynamics.

ESTIMATION OF FACTORS AND DFM PARAMETERS The key insight that makes high-dimensional DFM modeling practical is that if the number of series n is large, the factors can be estimated by suitable cross-sectional averaging. This is most easily seen in the special case of a single factor with a nonzero cross-sectional average value of the factor loadings. Let $\overline{X}_t$ denote the cross-sectional average of the variables at date t, $\overline{X}_t = n^{-1}\sum_{i=1}^{n} X_{it}$, and similarly let $\overline{\Lambda}$ and $\overline{e}_t$ respectively denote the cross-sectional average factor loading and the cross-sectional average of the idiosyncratic term. By equation 1, the cross-sectional average of the data satisfies $\overline{X}_t = \overline{\Lambda} F_t + \overline{e}_t$. But by assumption the idiosyncratic terms are only weakly correlated, so by the weak law of large numbers, $\overline{e}_t$ tends to zero as the number of series increases. Thus, when n is large, $\overline{X}_t$ consistently estimates $\overline{\Lambda} F_t$; that is, $\overline{X}_t$ estimates the factor up to scale and sign. In this special case, picking the arbitrary normalization $\overline{\Lambda} = 1$ yields the estimated factor time series, $\hat{F}_t$, and the individual factor loadings can be estimated by regressing each X_{it} on $\hat{F}_t$. If in fact there is a single-factor structure and n is sufficiently large, $\hat{F}_t$ estimates F_t precisely enough that $\hat{F}_t$ can be treated as data without a "generated regressor" problem (Bai and Ng 2006).

With multiple factors and general factor loadings, this simple cross-sectional averaging does not produce a consistent estimate of the factors, but the idea can be generalized using principal components analysis (Stock and Watson 2002). We use principal components here to estimate the factors; details are discussed in section I.D.

The principal components estimator of the factors consistently estimates F_t up to premultiplication by an arbitrary nonsingular $r \times r$ matrix (the analogue of $\overline{\Lambda}$ in the single-factor example); that is, the principal components estimator consistently estimates not the factors, but rather the space spanned by the factors when n and T are large. This means that the principal components estimator of F_t has a normalization problem, which is "solved" by the arbitrary restriction that $\Lambda'\Lambda = I_r$, the $r \times r$ identity matrix. This arbitrary normalization means that the individual factors do not have a direct economic interpretation (such as an "oil factor"). The analysis in

---

4. Equations 1 and 2 are the static form of the DFM, so called because the factors F_t enter with no leads or lags in equation 1. For a discussion of the relationship between the dynamic and the static forms of the DFM, see Stock and Watson (2011).

sections II and IV works with the reduced-form DFM in equations 1 and 2, so for that analysis this normalization is inconsequential. The analysis in section III requires identification of specific economic shocks, and our identification procedure is discussed there.

I.B. The Data and Preliminary Transformations

The data set consists of quarterly observations from 1959Q1 through 2011Q2 on 200 U.S. macroeconomic time series (vintage November 2011). The series are grouped into 13 categories: variables from the national income and product accounts (NIPA; 21 series); industrial production (13); employment and unemployment (46); housing starts (8); inventories, orders, and sales (8); prices (39); earnings and productivity (13); interest rates and spreads (18); money and credit (12); stock prices and wealth (11); housing prices (3); exchange rates (6); and other (2).

The series were subjected to a preliminary screen for outliers and then transformed as needed to induce stationarity. The transformation used depends on the category of the series. Real activity variables were transformed to quarterly growth rates (first differences of logs), prices and wages were transformed to quarterly changes of quarterly inflation (second differences of logs), interest rates were transformed to first differences, and interest rate spreads appear in levels. The 200 series and their transformations are listed in the online appendix.

I.C. Local Means and Detrending

All series were detrended to eliminate very low frequency variation. Specifically, after transforming the series to stationarity, we calculated the deviations of each series from a local mean estimated using a biweight kernel with a bandwidth of 100 quarters. These local means are approximately the same as those computed as the average of the transformed data over a centered moving window of ±30 quarters, except that the former are less noisy because they avoid the sharp cutoff of a moving window.[5] We refer to the local mean as the trend in the series, although it is important to note

5. Endpoints are handled by truncating the kernel and renormalizing the truncated weights to add to 1. This approach has the desirable feature that it makes no assumption about reversion to the local mean, in contrast to the mean reversion imposed by the standard approach of using a stationary time-series model to pad the series with forecasts and backcasts. We alternatively computed the local means using a Baxter-King (1999) high-pass filter with a pass band of periods with ≤ 200 quarters, and using the trend implied by a "local level" model (the sum of independent random walk and white noise with a ratio of disturbance standard deviations of 0.025), and obtained similar results. The weights for these different filters are given in the online appendix.

Figure 1. Growth Rates of GDP, Nonfarm Employment, Employee-Hours, and Labor Productivity, Quarterly and Trend, 1959Q1–2011Q2
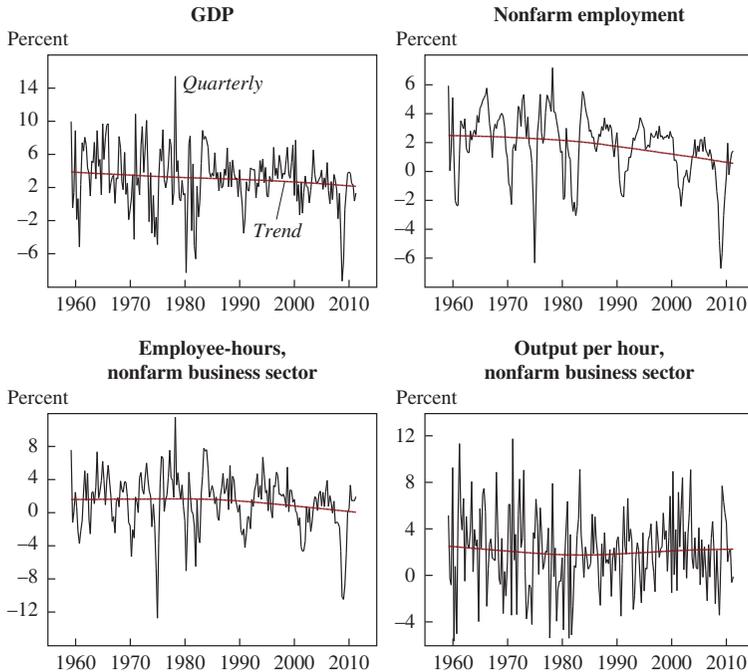
Sources: Data sources listed in the online appendix and authors' calculations.

that these are trends in transformed series; for example, for GDP the estimated trend is the local mean value of GDP growth.

For some series, the values of these trends (that is, local means) change substantially over the 1959–2011 period. Figure 1 plots the quarterly growth rates of GDP, employment, employee-hours, and labor productivity, along with their trends. We estimate the trend GDP growth rate to have fallen 1.2 percentage points, from 3.7 percent per year in 1965 to 2.5 percent per year in 2005,[6] and the trend annual employment growth

---

6. Our procedure produces a smooth but not necessarily monotonic trend. Kim and Eo (2012) model the trend decline in the growth rate of GDP as a single Markov switching break and estimate a decline of 0.7 percentage point over this period, less than our estimate of 1.2 percentage points. If the trend is in fact smoothly declining, one would expect their step-function approximation to estimate a smaller average decline than our local mean.

rate to have fallen by 1.5 percentage points, from 2.4 percent in 1965 to 0.9 percent in 2005. On the other hand, trend productivity (output per hour) has recovered from the productivity slowdown of the 1970s and 1980s and shows essentially no net change over this period. These trends are discussed further in section V.

I.D. Estimation Details

The data set contains both high-level aggregates and disaggregated components. To avoid double counting, in these cases only the disaggregated components were used to estimate the factors; for example, durables consumption, nondurables consumption, and services consumption were used to estimate the factors, but total consumption was not. Of the 200 series, 132 were used to estimate the factors; these are listed in the online appendix. No top-level macroeconomic aggregates (including GDP, consumption, investment, total employment, and the total unemployment rate) were used to estimate the factors.

Using these 132 series, we estimated the factor loadings Λ by principal components over 1959–2007Q3. These pre-2007Q4 factor loadings were then used to estimate the six linear combinations of X_t over the full 1959Q1–2011Q2 sample, which correspond to the estimated factors. For the 1959Q1–2007Q3 sample, these are the principal components of the factors; for the post-2007Q3 period, these are the pre-2007Q4 principal components factors, extended through the 2007Q4–2011Q2 sample. We refer to these six factors as the "old" factors because they are the factors for the "old" (pre-2007Q4) DFM, extended beyond 2007Q3.[7] These "old" factors are used throughout the paper; with the single exception of a sensitivity

---

7. Specifically, let $\tilde{X}_t$ denote the vector of 132 disaggregated time series used to estimate the factors, and let $\tilde{\Lambda}$ denote their corresponding factor loadings. These factor loadings $\tilde{\Lambda}$ are estimated by principal components using data on $\tilde{X}_t$ over 1959Q1–2007Q3 (modified for some series having missing observations; see the online appendix). Denote the resulting estimates of $\tilde{\Lambda}$ by $\hat{\tilde{\Lambda}}^{59-07}$, normalized so that $\hat{\tilde{\Lambda}}^{59-07\prime}\hat{\tilde{\Lambda}}^{59-07} = I$. The estimated "old" factors are computed using $\hat{\tilde{\Lambda}}^{59-07}$ as $\hat{F}_t^{59-07} = \hat{\tilde{\Lambda}}^{59-07\prime}\tilde{X}_t$, $t = 1959Q1, \ldots , 2011Q2$. The values of $\hat{F}_t^{59-07}$ post-2007Q4 are those of the "old" factors in the sense that they are based on the pre-2007Q4 linear combinations of $\tilde{X}_t$. The factor loadings for the remaining 68 series (the high-level aggregates) are obtained by regressing each series on $\hat{F}_t^{59-07}$ using data through 2007Q3; these estimates, combined with $\hat{\tilde{\Lambda}}^{59-07}$, yield the estimated "old" factor loadings, $\hat{\Lambda}^{59-07}$. The vector of common components of the full vector of time series associated with these "old" factors and "old" factor loadings is $\hat{\Lambda}^{59-07}\hat{F}_t^{59-07}$.

check in section II.B, the "old" DFM coefficients, estimated over 1959–2007Q3, are also used throughout.[8]

The DFM is estimated with six factors, a choice consistent with Bai-Ng (2002) tests for the number of factors, visual inspection of the scree plot, and the number of distinct structural shocks we examine in section IV.[9] As discussed below and shown in the online appendix, our main results are not very sensitive to varying the number of factors over a reasonable range.

II. A Structural Break in 2007Q4?

This section investigates the extent to which the 2007–09 recession exhibited new macrodynamics relative to the 1959Q1–2007Q3 experience. This analysis has three parts. First, we examine whether the factors in the 2007–09 recession were new or, alternatively, were combinations of "old" factors seen in previous recessions. Second, to the extent that at least some of the shocks have historical precedents, we examine whether these "old" factors have different dynamic impacts before 2007Q4 than in 2007Q4–2011. Third, we examine the volatility of these "old" factors over the recession. The analysis in this section uses the reduced-form factors and does not require identifying individual structural shocks.

II.A. *Post-2007 Simulation Using the Pre-2007 DFM*

We begin by considering the following experiment: suppose a forecaster in 2007Q3 had in hand our six-factor, 200-variable DFM estimated using data through 2007Q3 and was magically given a sneak preview of the six

---

8. The assumption that the factors and DFMs can be estimated with no breaks over the 1959Q1–2007Q3 period is only partially consistent with the empirical evidence. On the side of stability, in Stock and Watson (2009) we use a similar data set and find that the space spanned by the full-sample (no-break) factors spans the space of the factors estimated using pre- and post-1984 subsamples; against this, there we also find breaks in some factor loadings in 1984Q1. These apparently contradictory findings can be reconciled by the property of DFMs that the space spanned by the factors can be estimated consistently even if there is instability in Λ (Stock and Watson 2002, 2009, Bates and others 2012). These findings suggest that, in the present analysis, we can ignore the 1984Q1 break when estimating the factors; however, tests of coefficient stability might be sensitive to whether the comparison sample includes pre-1984Q1 data. We therefore consider a DFM with a break in 1984Q1 as a sensitivity check. Additional sensitivity checks, with breaks in 1984Q1, are reported in the online appendix.

9. The Bai-Ng (2002) IC_{P1} and IC_{P2} criteria select either three or four factors, depending on the sample period, whereas their IC_{P3} criterion selected 12 factors. The scree plot (the plot of the ordered eigenvalues of the sample covariance matrix of X_t) drops sharply to four or five factors and then declines slowly.

"old" factors (but only those six factors) from 2007Q4 through 2011Q2. Using the pre-2007Q4 model and the post-2007Q3 values of the old factors, this hypothetical forecaster computes predicted values for all 200 series in our data set. How well would these predicted values track the actuals over the recession and recovery? If there were an important new factor not seen in the 1959–2007Q3 data—say, a seventh, "financial crisis" factor—that factor would end up in the error term, and the fraction of the variation in the 200 macro variables explained by the "old" factors would be lower after 2007Q4 than before. Similarly, if the "old" factors had new effects—that is, if their factor loadings changed—then again the R^2s computed using the pre-2007Q4 factor loadings would drop.

The results of this exercise are summarized in figure 2 and in tables 1 and 2. Figure 2 plots the common components as predicted using the "old" model and factors (computed as described in footnote 7), along with actual values, for 24 selected time series from 2000Q1 through 2011Q2. For activity variables and inflation, the figure plots the 4-quarter growth rate (that is, the annualized 4-quarter average of quarterly growth) to smooth over quarterly measurement error, whereas for financial variables the figure plots quarterly changes or levels to provide a better picture of the financial market volatility of 2008–09.

Table 1 summarizes the patterns observed in figure 2 by reporting the R^2s of the common components of the 24 selected series, computed over various subsamples: two split-sample periods and the 15-quarter stretches following all postwar cyclical peaks. These R^2s are computed imposing a zero intercept and unit slope on the old model/old factors predicted values over the indicated subsample and thus cannot exceed 1 but can be negative.[10] The R^2s in table 1 all pertain to quarterly values, transformed as described in section I.B, whereas, as already noted, some plots in figure 2 are 4-quarter values.

The results in figure 2 and table 1 suggest that knowledge of the historical DFM and future values of the "old" factors explains most—for some series, nearly all—of the movements in most of the 200 macroeconomic time series. The predicted values in figure 2 capture the initially slow decline in early 2008, the sharp decline during 2008Q4–2009, the prolonged trough, and the muted recovery since 2010 in GDP, total consumption, nonresidential fixed investment, industrial production, employment,

---

10. Using the notation of footnote 7, let $\hat{e}_t = X_t - \hat{\Lambda}^{59-07}\hat{F}_t^{59-07}$ be the prediction error using the "old" model and factors. The subsample R^2 for series i is computed as $R^2 = 1 - \left(\sum_t \hat{e}_{it}^2\right)\Big/\left(\sum_t X_{it}^2\right)$ where the sums are computed over the column subsample.

Figure 2. Selected Macroeconomic Variables, Actual and Common Components, 2000Q1–2011Q2[a]

Four-quarter change in GDP

Percent

Four-quarter change in total consumption

Percent

Four-quarter change in services consumption

Four-quarter change in nonresidential fixed investment

Four-quarter change in industrial production (total)

Four-quarter change in industrial production (automobiles)

Four-quarter change in nonfarm employment

Four-quarter change in unemployment rate

Four-quarter change in short-term unemployment rate[c]

Four-quarter change in long-term unemployment rate[c]

Figure 2. Selected Macroeconomic Variables, Actual and Common Components, 2000Q1–2011Q2[a] (*Continued*)

Four-quarter change in housing starts

Percent

Four-quarter change in OFHEO house price index[d]

Percent

Four-quarter change in PCE inflation[e]

Four-quarter change in PCE gas and energy inflation[e]

Quarterly change in federal funds rate

Quarterly change in real monetary base

Quarterly change in commercial and industrial loans

TED spread[f,g]

Gilchrist-Zakrajšek spread[f,h]

Quarterly change in S&P 500 index

(*continued*)

Figure 2. Selected Macroeconomic Variables, Actual and Common Components, 2000Q1–2011Q2[a] (*Continued*)
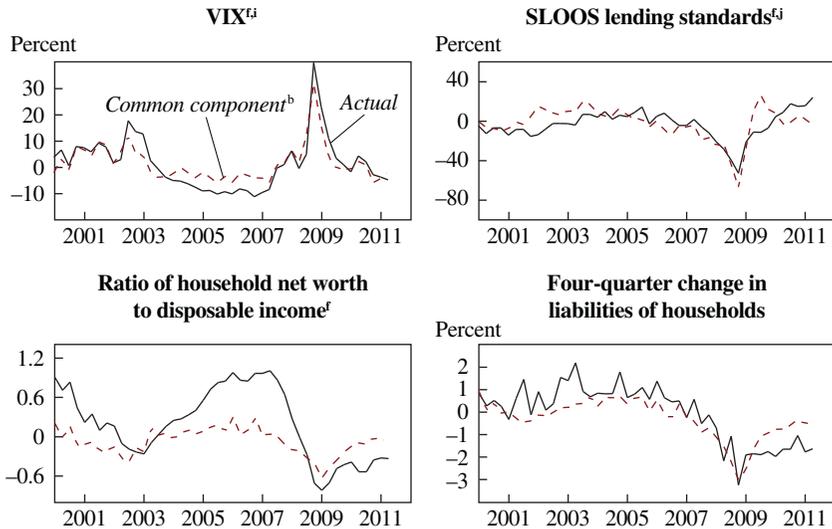
Sources: Authors' calculations; see the online appendix for details of the data sources.

a. For real activity variables and prices, which are in logarithms, the 4-quarter or quarterly changes are scaled to be percentage changes at an annual rate.

b. Computed using pre-2007Q4 coefficients and 2007Q4–2011 values of the "old" factors, derived from the 1959–2007Q3 DFM model as described in footnote 7 in the text.

c. Short-term and long-term unemployment refer to spells of <27 weeks and ≥27 weeks, respectively.

d. Reported values are adjusted for inflation. OFHEO = Office of Federal Housing Enterprise Oversight. Since 2008 the index has been published by the Federal Housing Finance Agency (FHFA) and is known as the FHFA house price index.

e. Four-quarter change in 400 times the change in log prices. PCE = personal consumption expenditures.

f. The series is the level of the indicator, not a change.

g. Difference between the 3-month London interbank offer rate and the 3-month Treasury interest rate.

h. Excess bond premium spread from Gilchrist and Zakrajšek (forthcoming).

i. Chicago Board of Options Exchange Market Volatility Index.

j. Federal Reserve Senior Loan Officer Opinion Survey measure of lending standards (net percentage of domestic respondents tightening standards for commercial and industrial loans).

and the unemployment rate. The pre-2007Q4 model and historical factors predict the prolonged, accelerating decline of housing starts, although the anemic recovery of housing is slightly overpredicted. Given these factors, there are no major surprises in overall inflation or even energy price inflation. The historical factors even explain the general pattern of interest rate spreads (the TED spread and the Gilchrist-Zakrajšek excess bond premium spread, from Gilchrist and Zakrajšek forthcoming), the bear market in stocks, and the sharp rise in uncertainty as measured by the VIX, and even

Table 1. Subsample R^2's of Regressions of Common Components on Actual Data for Selected Macroeconomic Variables[a]

Series	1959–2007Q3	1984–2007Q3	Computed over 15 quarters starting at indicated cyclical peak							
			1960Q2	1969Q4	1973Q4	1980Q1	1981Q3	1990Q3	2001Q1	2007Q4
GDP	0.73	0.62	0.81	0.79	0.83	0.79	0.85	0.73	0.64	0.64
Consumption (total)	0.63	0.45	0.59	0.77	0.66	0.82	0.72	0.74	0.02	0.57
Consumption (services)	0.35	0.22	0.28	0.46	-0.06	0.62	0.47	0.47	-0.25	0.84
Nonresidential fixed investment	0.63	0.57	0.64	0.77	0.90	0.58	0.57	0.50	0.69	0.86
Industrial production (total)	0.87	0.79	0.90	0.86	0.94	0.93	0.94	0.89	0.76	0.95
Industrial production (automobiles)	0.58	0.29	0.59	0.61	0.70	0.64	0.39	0.66	0.13	0.57
Nonfarm employment	0.92	0.91	0.93	0.94	0.93	0.95	0.98	0.93	0.91	0.96
Unemployment rate	0.84	0.75	0.77	0.88	0.95	0.91	0.88	0.76	0.79	0.89
Short-term unemployment rate	0.82	0.70	0.81	0.87	0.90	0.88	0.84	0.81	0.75	0.78
Long-term unemployment rate	0.61	0.58	0.55	0.62	0.68	0.74	0.71	0.44	0.51	0.64
Housing starts	0.59	0.39	0.27	0.59	0.83	0.79	0.76	0.62	0.29	0.53
OFHEO house price index	0.49	0.44	n.a.	n.a.	n.a.	0.59	-0.01	0.60	0.42	0.57
PCE inflation	0.51	0.54	0.25	0.70	0.66	0.44	0.40	0.57	0.56	0.82
PCE gas and energy inflation	0.37	0.48	-0.42	-1.69	-0.63	0.25	0.23	0.55	0.48	0.71
Federal funds rate	0.44	0.34	-0.01	0.47	0.57	0.48	0.41	0.72	0.14	-1.51
Real monetary base	0.16	0.09	0.18	0.39	0.22	0.65	0.59	-0.18	-0.35	-0.03
Real commercial and industrial loans	0.44	0.54	0.48	0.47	0.55	-1.42	-1.07	0.70	0.62	0.46
TED spread	0.54	0.02	n.a.	n.a.	0.78	0.74	0.70	-0.05	-0.04	0.76
Gilchrist-Zakrajšek spread	0.48	0.48	n.a.	n.a.	0.89	0.11	0.49	0.12	0.65	n.a.
S&P 500 index	0.72	0.69	0.75	0.68	0.89	0.62	0.65	0.37	0.81	0.89
VIX	0.47	0.50	n.a.	0.38	0.74	-0.78	-0.69	0.40	0.75	0.89
SLOOS lending standards	0.60	0.18	n.a.	n.a.	0.65	0.79	0.67	0.55	-1.49	0.47
Household net worth–disposable income ratio	0.16	0.14	-0.18	0.67	0.48	0.78	0.06	0.28	-0.46	0.51
Household liabilities	0.57	0.48	-0.19	0.73	0.90	0.76	0.70	0.79	0.18	0.77

Source: Authors' calculations; see the notes to figure 1 and the online appendix for further details.

a. Common components are calculated using the six static factors from the DFM estimated using quarterly data over 1959Q1–2007Q3; predicted values are the "old model/old factors" values computed as described in footnote 7 in the text. All entries are 1 minus the ratio of the sum of squared prediction errors to the sum of squares of the observed variable (see footnote 10). "n.a." indicates that the regression was not performed because data for that series were not available for the indicated period. All series are transformed and detrended as described in sections I.B and I.C and in the online appendix

Table 2. Subsample R^2s of the Common Component Regressions by Category[a]

				R^2								
				Computed over 15 quarters starting at indicated cyclical peak								
Series category	*N*	*1959– 2007Q3*	*1984– 2007Q3*	*1960Q2*	*1969Q4*	*1973Q4*	*1980Q1*	*1981Q3*	*1990Q3*	*2001Q1*	*2007Q4*
National income and product accounts	21	0.56	0.43	0.56	0.46	0.62	0.66	0.59	0.64	0.49	0.65
Industrial production	13	0.72	0.60	0.78	0.77	0.86	0.86	0.80	0.66	0.61	0.80
Employment and unemployment	46	0.62	0.50	0.64	0.68	0.76	0.76	0.81	0.61	0.63	0.78
Housing starts	8	0.37	0.21	0.09	0.26	0.54	0.46	0.54	0.44	−0.16	0.27
Inventories, orders, and sales	8	0.54	0.35	0.39	0.69	0.72	0.73	0.68	0.66	0.43	0.64
Prices	39	0.15	0.05	0.00	0.15	0.37	0.18	0.08	0.03	0.14	0.13
Earnings and productivity	13	0.37	0.29	0.52	0.38	0.35	0.30	−0.03	0.32	0.36	−0.11
Interest rates	18	0.40	0.30	−0.07	0.39	0.30	0.50	0.44	0.29	0.07	−0.50
Money and credit	12	0.44	0.26	0.22	0.47	0.55	0.65	0.63	0.40	0.06	−0.37
Stock prices and wealth	11	0.47	0.52	0.00	0.67	0.74	0.62	0.50	0.37	0.31	0.77
Housing prices	3	0.67	0.67	n.a.	n.a.	n.a.	0.59	−0.01	0.60	0.72	0.57
Exchange rates	6	0.56	0.66	−2.64	0.47	0.64	0.48	0.66	0.75	0.72	0.60
Other	2	0.42	0.42	−0.40	0.15	0.87	0.89	0.89	0.48	−0.60	0.31

Source: Authors' calculations.

a. Based on the six-factor DFM estimated over 1959Q1–2007Q3. Entries are the median of the R^2s of the common component, computed for the series in the row category over the indicated period, where the R^2 is computed as described in the notes to table 1. "n.a." indicates that the regression was not performed because data for that series were not available for the indicated period.

the sharp decline and recovery of lending standards reported in the Federal Reserve's Senior Loan Officer Opinion Survey. The DFM correctly predicts the decline in commercial and industrial loans during the early part of the recession, although it underpredicts the depth of their contraction and their long delay in recovering.[11] These qualitative impressions from figure 2 are confirmed quantitatively by the R^2s in table 1. For these series, the post-2007Q3 R^2s are well within the range of R^2s for previous recessions. The post-2007Q3 R^2 for GDP growth is somewhat lower than in previous episodes because the DFM misses some high-frequency variation, but as seen from the first panel of figure 2, the year-over-year match is very strong. On the other hand, for some series (among those in table 1, consumption of services, personal consumption expenditures inflation, and the VIX), the post-2007Q4 R^2s are substantially greater than their historical averages. One interpretation of this improved fit during 2007Q4–2011 is that the movements in the common component of these series, computed using the pre-2007Q4 factors, were so large during this recession that the fraction of the variation it explains increased.

A few series are less well explained by the historical factors. Most notably, the model predicts a larger decline in the federal funds rate than occurred, but this is unsurprising because the model is linear and lacks a zero lower bound; we return to this point below. The model also confirms that the Federal Reserve's expansion of reserves was unprecedented. Although the historical factors predict home prices in 2007Q4–2011 as well as in previous recessions, they do not fully explain the boom in home prices in 2004–06, and they slightly underpredict the speed of their crash.

Table 2 summarizes the subsample R^2s for all 200 series, by series category (the online appendix reports results corresponding to those in figure 2 and table 1 for all series). For most categories the median R^2 over the period since 2007Q4 is comparable to or greater than that in previous recessions and recoveries. The only categories for which the predicted and actual paths diverge systematically are earnings and productivity, interest rates, and money and credit. The divergence in interest rates is due mainly to problems relating to the zero lower bound, not to failures to match liquidity spikes in the spreads, and the divergence in money and credit is associated

---

11. Giannone, Lenza, and Reichlin (2012) examine the stability of the relationship between various types of loans and macroeconomic indicators in the euro zone during and after the crisis, relative to a precrisis benchmark; they find no surprising behavior of loans to nonfinancial corporations, conditional on aggregate activity, although there are departures from historical patterns for household loans.

with the unprecedented expansion of monetary aggregates. Closer inspection of the divergence in earnings and productivity suggests that it does not reflect breaks associated with this recession compared with the two other post-1984 recessions.[12]

II.B. Tests for a Break in the Factor Loadings, 2007Q4–2011

The results in the previous subsection suggest that the DFM did not suffer a structural break or regime shift in the 2007–09 recession. We now turn to two tests of this hypothesis.

The first test is of the hypothesis that the factor loadings are constant, against the alternative that they suffered a break in 2007Q4–2011. We do this using Donald Andrews' (2003) test for end-of-sample instability.[13] As discussed above, there is evidence of a break in 1984Q1 in a substantial fraction of the factor loadings. We therefore consider two versions of the Andrews (2003) test, one testing the hypothesis of stability of a break in 2007Q4 relative to the 1959Q1–2007Q3 values of the loadings, and the other testing for a break in 2007Q4 relative to the values of the loadings over 1984Q1–2007Q3.

Rejection rates of this test for a break in 2007Q4 at the 5 percent level are summarized by series category in table 3. When the post-2007 values are compared with the full-sample factor loadings, the hypothesis is rejected at the 5 percent level for 15 percent of all series, and for 12 percent when the comparison is with the 1984Q1–2007Q3 factor loadings (final column of table 3). This slightly higher rejection rate for the tests against the full pre-2007Q4 sample is consistent with a break in the factor loadings in 1984Q1 as found in Stock and Watson (2009).

When evaluated against the 1984Q1–2007Q3 loadings, all but a handful of rejections are concentrated in three areas: commodity and materials producer price inflation indexes, the durational composition of unemployment, and monetary aggregates. Some examples are shown in figure 2 (see the panels for long-term unemployment and the monetary base). The small number of rejections provides little evidence of a systematic or widespread

---

12. The negative quarterly R^2s for output per hour reflect a timing mismatch, and 4-quarter growth in productivity is well predicted. The predicted values for average hourly earnings growth change from procyclical to countercyclical in the mid-1980s, and the negative R^2 reflects this apparent instability in the factor loadings in 1984, not something special to the 2007–09 recession.

13. The Andrews (2003) test is based on an analogue of the usual (homoskedasticity-only) Chow break-test statistic, with a p value that is computed by subsampling.

Table 3. Tests of Absence of a Break in Factor Loadings

Series category	N	Percent of series in the category for which the hypothesis of a break at 2007Q4 is rejected at the 5 percent level[a]	
		1959–2007Q3	*1984Q1–2007Q3*
National income and product accounts	21	0	0
Industrial production	13	0	0
Employment and unemployment	46	15	15
Housing starts	8	25	13
Inventories, orders, and sales	8	13	13
Prices	39	26	23
Earnings and productivity	13	15	8
Interest rates	18	0	11
Money and credit	12	42	17
Stock prices and wealth	11	18	0
Housing prices	3	33	33
Exchange rates	6	0	0
Other	2	0	0

Source: Authors' calculations.

a. The Andrews (2003) end-of-sample stability test is used to test the hypothesis of stability of the factor loadings. The statistic tests the null hypothesis of constant factor loadings against the alternative of a break in the final 15 quarters (2007Q4–2011Q2) relative to the value of the factor loading estimated over the indicated period.

break in the factor loadings in 2007Q4, relative to their Great Moderation values.

As a second test, we examine evidence for a new factor by testing whether the idiosyncratic disturbances, computed relative to the pre-2007Q4 factors, show unusual evidence of common structure in the current recession. Specifically, we used the pre-2007Q4 factors to compute the vector of idiosyncratic disturbances for the 8 quarters starting with 2007Q4 (in the notation of footnote 7, this vector is $X_t - \hat{\Lambda}^{59-07} \hat{F}_t^{59-07}$). The sample second-moment matrix of these disturbances has rank 8, and the ratio of the first eigenvalue of this matrix to the sum of all eight nonzero eigenvalues is a measure of the correlation among the idiosyncratic disturbances during these 8 quarters. A new common factor would produce an unusually large value of this ratio. It turns out that this eigenvalue ratio following 2007Q3 is less than it was during the recessions that began in 1960Q2 and 1973Q4. The p value testing the hypothesis that this ratio is the same as its pre-2007Q4 mean, computed by subsampling consecutive 8-quarter periods, is 0.59. Modifying the subsampling test to examine the 15 quarters starting

with 2007Q4 yields a *p* value of 0.90. This test therefore provides no evidence of a missing factor.

II.C. *Increased Variance of Factors*

The findings of sections II.A and II.B suggest that the severity of the 2007–09 recession was associated with large unexpected movements in the factors, not with some new factor or with changes in macroeconomic dynamics (changes in coefficients). Indeed, the factors were highly volatile over this period. Table 4 summarizes the standard deviations of selected variables over the pre–1984 period, the Great Moderation period, and the post-2004 period, along with the standard deviations of their factor components computed using pre-2007Q3 coefficients and the "old" factors. For these (and other) macroeconomic aggregates, volatility since 2004 has returned to or exceeds pre–Great Moderation levels. As the right-hand panel of table 4 shows, this increased volatility is associated with increased volatility of the factor components, which (because the coefficients are constant) derives from increased volatility of the factors themselves.

Table 5 takes a closer look at the factor innovations over this period. Because the factors are identified by the arbitrary normalization of principal components analysis, the innovations to individual factors are hard to interpret. Table 5 therefore examines linear combinations of the factor innovations determined by the factor loadings for various macroeconomic variables, so that the table entries are the innovations in the common component, by series, by quarter, from 2007Q1 through 2011Q2, reported in standard deviation units.[14] Among the series in table 5, the factor component of oil prices experienced a moderate positive standardized innovation in 2007Q1, then a large positive standardized innovation in 2008Q2 (1.7 and 3.4 standard deviations, respectively), and the TED spread, the VIX, and housing starts experienced large innovations in 2008Q3, then extremely large (approximately 8 standard deviations) innovations in 2008Q4. Oil prices experienced a very large negative factor innovation in 2008Q4, then large positive innovations in the next three quarters. Throughout 2007Q4–2009Q1, the factor component innovations for the real variables were moderate by comparison and were generally within the range of pre-2007Q4 experience. By 2009Q4, all the innovations had

---

14. Rewrite the factor VAR (equation 1) as $F_t = \tilde{\Phi}(L)F_{t-1} + \eta_t$, so, from equation 2, $X_t = \Lambda F_t + e_t = \Lambda\tilde{\Phi}(L)F_{t-1} + \Lambda\eta_t + e_t$. Then $\Lambda\tilde{\Phi}(L)F_{t-1}$ is the contribution of the past factors and $\Lambda\eta_t$ is the innovation in the common component. The innovations in table 5 are the residuals from a four-lag VAR estimated using the "old" factors over 1959–2011Q2.

Table 4. Standard Deviations of Four-Quarter Growth Rates of Major Activity Variables[a]

Series	Standard deviation of series			Standard deviation of factor component		
	1959–83	1984–2004	2005–11[b]	1959–83	1984–2004	2005–11[b]
GDP	2.6	1.6	2.8	2.6	1.4	2.8
Consumption	2.1	1.3	2.3	2.1	1.2	2.4
Investment	11.4	8.5	15.3	10.5	6.8	11.4
Industrial production (total)	5.2	3.1	6.4	4.9	3.2	5.9
Nonfarm employment	2.0	1.4	2.3	1.9	1.3	2.5
Change in unemployment rate[c]	1.1	0.8	1.5	1.1	0.7	1.4

a. Entries are standard deviations of 4-quarter detrended growth rates, except where noted otherwise, over the indicated period. The first three columns report standard deviations of the row series, and the second three columns report standard deviations of the factor component of the row series.

b. Calculations go through the final quarter in the data set, 2011Q2.

c. Entries refer to the 4-quarter detrended change in the unemployment rate.

Table 5. Innovations to Factor Components of Selected Series, 2007Q1–2011Q2

Standard deviation units

Quarter	GDP	Total consumption	Investment	Employment	Productivity	Housing starts	Oil price	Federal funds rate	TED spread	VIX	Household wealth
2007Q1	−0.9	−1.3	−0.4	−0.7	−1.2	0.2	1.7	0.3	0.0	−0.6	0.0
2007Q2	0.3	−0.2	0.5	−0.1	0.2	0.8	1.1	0.5	−0.9	−1.4	0.8
2007Q3	−0.3	−0.8	0.0	−0.7	0.0	−0.7	0.3	−0.6	0.7	1.2	−1.0
2007Q4	−0.3	−1.3	0.1	0.3	−0.7	−1.3	1.3	0.4	0.4	0.5	−0.9
2008Q1	−0.3	−0.7	0.1	0.2	−0.4	−1.3	0.2	−0.1	1.4	2.2	−2.0
2008Q2	−1.4	−2.1	−0.4	−1.1	−2.1	1.0	3.4	0.5	0.1	−0.4	−0.8
2008Q3	−1.7	−1.7	−1.0	−0.7	−1.4	−3.5	−0.6	0.2	3.9	2.9	−2.6
2008Q4	1.0	2.1	−0.1	−0.4	4.6	−8.3	−10.3	−2.5	7.7	8.3	−4.1
2009Q1	0.3	−2.7	2.3	0.9	−0.3	−4.7	2.5	3.5	4.0	1.4	−3.3
2009Q2	2.9	1.8	3.3	3.8	0.7	2.9	2.8	3.8	−3.0	−3.4	1.2
2009Q3	1.6	0.4	1.9	2.3	−0.6	4.8	5.0	1.5	−5.2	−3.5	1.5
2009Q4	−1.2	−0.9	−2.0	−2.1	−0.2	0.1	−0.4	−2.1	−1.7	−2.1	2.7
2010Q1	0.3	−0.1	0.6	0.5	0.0	−0.5	0.3	1.2	0.9	0.0	−0.6
2010Q2	0.7	0.3	0.7	0.3	1.3	−2.4	−1.8	0.3	2.1	1.6	−1.2
2010Q3	0.8	−0.3	1.1	0.4	0.9	−1.7	0.0	0.9	1.0	0.3	−0.7
2010Q4	0.3	−0.6	0.6	0.0	0.0	0.6	1.7	0.8	−1.0	−1.8	0.8
2011Q1	0.4	−0.6	1.1	0.5	−0.6	1.5	2.8	1.2	−0.9	−0.8	−0.5
2011Q2	−0.9	−0.8	−1.1	−1.3	−0.6	0.5	0.4	−1.5	−0.7	0.1	0.3

Sources: Authors' calculations.

a. Entries are the standardized innovations in the factor component of each series, computed relative to the six factors; series are standardized by dividing by the standard deviation of the 1959–2007Q3 factor component innovations for that series. Standardized innovations equal to or exceeding 3 in absolute value are italicized.

returned to their normal range, which is consistent both with the large economic shocks having passed and with the pre-2007Q4 model coefficients continuing to describe the macrodynamics.

The picture of the recession that emerges from table 5 is one of increases in oil prices through the first part of the recession, followed in the fall of 2008 by financial sector volatility, a housing construction crash, heightened uncertainty, and a sharp unexpected drop in wealth. Notably, there are few large surprise movements in the common components of the real variables, given the factors through the previous quarter.

II.D. Discussion

The results of this section suggest three main findings. First, there is little evidence of a new factor associated with the 2007–09 recession and its aftermath; rather, the factors associated with that recession are those associated with previous recessions and with economic fluctuations more generally from 1959 through 2007Q3. Second, for most of the series in our data set and in particular for the main activity measures, the response to these "old" factors seems to have been the same after 2007Q4 as before. Third, there were large innovations in these "old" factors during the recession, especially in the fall of 2008.

We believe that the most natural interpretation of these three findings is that the 2007–09 recession was the result of one or more large shocks, that these shocks were simply larger versions of ones that had been seen before, and that the response of macroeconomic variables to these shocks was almost entirely in line with historical experience. The few series for which behavior departed from historical patterns have straightforward explanations; in particular, the DFM predicts negative interest rates because it does not impose a zero lower bound, and the DFM does not predict the Federal Reserve's quantitative easing.

This interpretation comes with caveats. First, the stability tests in section II.B are based on 15 post-2007Q3 observations, so their power could be low; however, the plots in figure 2 and the R^2s in tables 1 and 2 provide little reason to suspect systematic instability that is missed by the formal tests.

Second, although the results concern the factors and their innovations, our interpretation shifts from factor innovations to shocks. A new shock that induced a new pattern of macrodynamics would surface in our DFM as a new factor, but we find no evidence of a missing or new factor. However, the possibility remains that there was a new shock in 2007Q4 that has the same effect on the factors as previously observed shocks. Indeed,

at some level this must be so: the Lehman collapse was unprecedented and the "Lehman shock" was new, and so were the TARP, quantitative easing, the auto bailout, and the other extraordinary events of this recession. Our point is that although all these particulars were new, their dynamic effect on the economy was not.

III. Structural Shocks: Identification and Contribution to the 2007–09 Recession

The analysis of section II suggests that the shocks precipitating the 2007–09 recession were simply larger versions of shocks experienced by the economy over the previous five decades. We now turn to the task of identifying those shocks and quantifying their impact, starting with our general approach to identification.

III.A. DFM Shock Identification Using Instrumental Variables

The identification problem in structural VAR analysis is how to go from the moving-average representation in terms of the innovations (the one-step-ahead forecast errors of the variables in the VAR) to the moving-average representation in terms of the structural shocks, which is the impulse response function with respect to a unit increase in the structural shocks. This is typically done by first assuming that the innovations can be expressed as linear combinations of the structural shocks, then by imposing economic restrictions that permit identification of the coefficients of those linear combinations. Those coefficients in turn identify the shocks and the impulse response function of the observed variables with respect to the shocks. This approach can be used to identify all the shocks, a subset of the shocks, or a single shock.

Most identification schemes for structural VAR analysis have an instrumental variables interpretation. When the economic restrictions take the form of exclusion restrictions on the impulse response function (shock A does or does not affect variable B within a quarter; shock C does or does not have a long-run effect on variable D), the restrictions turn certain linear combinations of the innovations into instrumental variables that in turn identify the structural impulse response functions. We refer to such instruments as "internal" instruments because they are linear combinations of the innovations in variables included in the VAR. An alternative method, pioneered by Christina Romer and David Romer (1989), is to use information from outside the VAR to construct exogenous components of specific shocks directly. These exogenous components are typically treated

as exogenous shocks; however, technically they are instrumental variables for the shocks: they are not the full shock series, but rather measure (typically with error) an exogenous component of the shock, so that the constructed series is correlated with the shock of interest but not with other shocks. We refer to these constructed series as "external" instruments, because they use information external to the VAR for identification. For example, one of our external instruments for the monetary policy shock is the Romer and Romer (2004) monetary shock series; there this series was treated directly as a monetary policy shock, whereas here it is taken to be correlated with the monetary policy shock and uncorrelated with all other structural shocks. More generally, in a structural VAR, an external instrument is a variable used for identification that is not itself included in the VAR; in a structural DFM, an external instrument is a variable used for identification that is not itself a factor. With one exception—a productivity shock instrument identified by a long-run exclusion restriction as in work by Jordi Galí (1999), discussed below—all the instruments used in this paper are external instruments.

IDENTIFICATION AND INFERENCE USING EXTERNAL INSTRUMENTS The basic idea of structural VAR identification with external instruments is that the structural shock is identified as the predicted value in the population regression of the instrument, say, Z_t, on the VAR innovations η_t.[15] For this result to hold, the instrument needs to be valid; that is, it must be relevant (correlated with the structural shock of interest) and exogenous (uncorrelated with all other structural shocks), and the structural shocks must be uncorrelated. We now summarize the math of this identification argument for the case of a single instrument, which is the relevant case for this paper because we estimate shocks using one instrument at a time. This discussion is written in terms of structural DFMs, but the argument applies directly to structural VARs with the interpretation that η_t are the reduced-form VAR innovations. For technical details, the extension to multiple instruments, system and subsystem estimation, and inference with weak and strong instruments, see Montiel Olea, Stock, and Watson (2012).

15. The approach to structural VAR identification laid out here, including the estimator in the just-identified case, was originally presented in Stock and Watson (2008). This approach was also developed independently in Mertens and Ravn (2012), of which we became aware after presenting the conference draft of this paper. The idea of using constructed exogenous shocks (what we call external instruments) as instruments in structural VARs dates at least to Hamilton (2003); also see Kilian (2008a, 2008b).

As is standard in the structural VAR literature, we assume that the r innovations η_t are linear combinations of r structural shocks ε_t, so that

$$
(3) \qquad \eta_t = H\varepsilon_t = \begin{bmatrix} H_1 & \cdots & H_r \end{bmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{rt} \end{pmatrix},
$$

where H_1 is the first column of H, ε_{1t} is the first structural shock, and so forth. Thus $\Sigma_{\eta\eta} = H\Sigma_{\varepsilon\varepsilon}H'$, where $\Sigma_{\eta\eta} = E(\eta_t\eta_t')$ and $\Sigma_{\varepsilon\varepsilon} = E(\varepsilon_t\varepsilon_t')$. We also assume, as is standard in the structural VAR literature, that the system described in equation 3 is invertible, so that the structural shocks can be expressed as linear combinations of the innovations:

$$
(4) \qquad\qquad \varepsilon_t = H^{-1}\eta_t.
$$

A key object of interest in structural VAR/DFM analysis is the impulse response function with respect to a structural shock. From equations 2 and 3, we have that $F_t = \Phi(L)^{-1}H\varepsilon_t$, which, when substituted into equation 1, yields

$$
(5) \qquad\qquad X_t = \Lambda\Phi(L)^{-1}H\varepsilon_t + e_t.
$$

The impulse response function of X_t with respect to the ith structural shock thus is $\Lambda\Phi(L)^{-1}H_i$. As discussed in section I, Λ and $\Phi(L)$ are identified from the reduced form, so it remains only to identify H_i.

We consider the problem of identifying the effect of a single shock, which for convenience we take to be the first shock ε_{1t}, using the single instrumental variable Z_t. The instrument and the shocks are assumed to satisfy three conditions:

$$
\begin{aligned}
&\text{(i)}\, E(\varepsilon_{1t}Z_t) = \alpha \neq 0\,(\text{relevance}) \\
(6) \quad &\text{(ii)}\, E(\varepsilon_{jt}Z_t) = 0,\, j = 2, \ldots, r\,(\text{exogeneity}) \\
&\text{(iii)}\, \Sigma_{\varepsilon\varepsilon} = D = \mathrm{diag}(\sigma_{\varepsilon_1}^2, \ldots, \sigma_{\varepsilon_r}^2)\,(\text{uncorrelated shocks}),
\end{aligned}
$$

where D in condition (iii) is an $r \times r$ diagonal matrix. Condition (i) says that Z_t is correlated with the shock of interest, ε_{1t}; that is, Z_t is a relevant instrument. Condition (ii) says that Z_t is uncorrelated with the other structural shocks. By conditions (i) and (ii), Z_t is correlated with η_t only because it is

correlated with ε_{1t}. Condition (iii) is the standard structural VAR assumption that the structural shocks are uncorrelated. Condition (iii) does not fix the shock variance, and normalization of the shocks is discussed below.

Conditions (i) and (ii) imply that

$$(7) \quad E(\eta_t Z_t) = E(H\varepsilon_t Z_t) = [H_1 \quad \cdots \quad H_r] \begin{pmatrix} E(\varepsilon_{1t} Z_t) \\ \vdots \\ E(\varepsilon_{rt} Z_t) \end{pmatrix} = H_1\alpha,$$

where the first equality follows from equation 3 and the final equality from conditions (i) and (ii). The instrument Z_t thus identifies H_1 up to scale and sign.

The shock ε_{1t} is identified (up to scale and sign) by further imposing condition (iii), which implies that $\Sigma_{\eta\eta} = HDH'$. Define Π to be the matrix of coefficients of the population regression of Z_t on η_t. Then, under conditions (i) through (iii),

$$(8) \qquad \Pi\eta_t = E(Z_t\eta_t')\Sigma_{\eta\eta}^{-1}\eta_t = \alpha H_1'(HDH')^{-1}\eta_t$$
$$= \alpha(H_1'H'^{-1})D^{-1}(H^{-1}\eta_t) = (\alpha/\sigma_{\varepsilon 1}^2)\varepsilon_{1t},$$

where the second equality follows from equation 7 and the final equality follows from equation 4 and $H^{-1}H_1 = e_1$, where $e_1 = (1, 0, \ldots, 0)'$ so that $\alpha(H_1'H'^{-1})D^{-1} = (\alpha/\sigma_{\varepsilon 1}^2)e_1'$.

Equation 8 displays the result anticipated in the opening sentence of this subsection: the shock identified using the instruments Z_t is the predicted value from the population regression of Z_t on the innovations η_t, that is, $\Pi\eta_t$, up to scale and sign. Additional intuition for this result is as follows. Suppose one observed ε_t, so that one could regress Z_t on ε_t; then, by condition (i) the population coefficient on ε_{1t} would be nonzero, whereas by conditions (ii) and (iii) the coefficients on the other ε_ts would be zero, so the predicted value would be ε_{1t} up to scale and sign. But by equations 3 and 4, the projection of Z_t on η_t has the same predicted value as the regression of Z_t on ε_t, so the predicted value from the population regression of Z_t on η_t is ε_{1t} (up to scale and sign).

The scale and sign of ε_{1t} and H_1 are set by normalizing the shock to have a unit impact on a given variable; for example, an oil price shock is normalized so that a 1-unit positive shock increases the (log) oil price by 1 unit.

ESTIMATION AND TESTS OF OVERIDENTIFYING RESTRICTIONS The structural shock is estimated using the sample analogue of equation 8; that is, $\hat{\varepsilon}_{1t}$ is

computed as the predicted value of the sample regression of Z_t on $\hat{\eta}_t$, where $\hat{\eta}_t$ is the vector of residuals from the reduced-form VAR estimated using $\hat{F}_t$. If Z_t is available only for a subperiod, the coefficients of this regression are used to compute the predicted values for the span for which $\hat{\eta}_t$ is available but Z_t is not. All subsequent calculations of interest here (decompositions, correlations, and so forth) are made using $\hat{\epsilon}_{1t}$.

CORRELATIONS AMONG IDENTIFIED SHOCKS Suppose one has two instruments that purportedly identify different shocks. If both instruments are valid, then in the population these identified shocks will be uncorrelated. But the population projection (equation 8) does not impose that the two shocks be uncorrelated; in fact, if one or both instruments are not valid, then in general the two shocks will be correlated. Similarly, two valid instruments that identify the same shock will produce identified shocks that are perfectly correlated in the population. The sample correlation between two estimated shocks therefore provides insight into the joint validity of the two instruments. Note that in general the correlation between the two identified shocks differs from the correlation between the two instruments themselves.

III.B. Instruments

We now turn to a discussion of the 18 instruments we use to identify structural shocks. We consider six structural shocks: to oil prices, monetary policy, productivity, uncertainty, liquidity and financial risk, and fiscal policy. Although this list is not exhaustive, these shocks feature prominently in discussions of the crisis and recovery, and each has a substantial literature upon which we can draw for its identification. The instruments are summarized here; specific sources and calculation details are provided in the online appendix.

OIL SHOCK We use three external instruments. The Hamilton (2003) oil shock is a quarterly version of James Hamilton's (1996) monthly net oil price increase, constructed over a 3-year window as in Hamilton (2003) as the percentage amount by which the oil price in a quarter exceeds the previous peak over the past 3 years (constructed from the producer price index for oil, available for 1960Q1–2011Q4). The Kilian (2008a) oil shock is Lutz Kilian's OPEC production shortfall stemming from wars and civil strife (1971Q1–2004Q3). The Ramey-Vine (2010) instrument is the residual from a regression of adjusted gasoline prices on various lagged macroeconomic variables as described in Valerie Ramey and Daniel Vine (2010), which we recomputed using the most recent data vintage. See Kilian (2008a, 2008b) and Hamilton (2009, 2010) for discussions of various oil shock measures.

MONETARY POLICY SHOCK We use four external instruments. The first is the Romer and Romer (2004) monetary policy shock, which they computed as the residual of a constructed Federal Reserve monetary intentions measure regressed on internal Fed forecasts (quarterly sums of their monthly variable, de-meaned, 1969Q1–1996Q4). The second is the shock to the monetary policy reaction function in Frank Smets and Raf Wouters' (2007) dynamic stochastic general equilibrium (DSGE) model, as recomputed by Robert King and Watson (2012; 1959Q1–2004Q4). The third is the monetary policy shock from Christopher Sims and Tao Zha's (2006) structural VAR allowing for shifts in shock variances but constant VAR coefficients (quarterly average of their monthly money shock, 1960Q1–2003Q4). The final instrument is the "target" factor of Refet Gürkaynak, Brian Sack, and Eric Swanson (2005), which measures surprise changes in the target federal funds rate (quarterly sums of daily data, 1990Q1–2004Q4).[16]

PRODUCTIVITY SHOCK We use one internal and two external instruments. The first external instrument is the series of Susanto Basu, John Fernald, and Miles Kimball (2006; Fernald 2009) on quarterly total factor productivity adjusted for variations in factor utilization, as updated by Fernald (1959Q1–2011Q2). The second external instrument is the productivity shock in the Smets-Wouters (2007) DSGE model, as recomputed by King and Watson (2012; 1959Q1–2004Q4). The internal instrument is constructed using Galí's (1999) identification scheme and is the permanent shock to the factor component of output per hour in nonfarm businesses. In DFM notation, let λ_{OPH}' denote the row of Λ corresponding to output per hour; then this internal instrument is $\lambda_{OPH}'\Phi(1)^{-1}\eta_t$ (1959Q1–2011Q2). Galí's (1999) identification scheme is controversial and has generated a large literature; see Karel Mertens and Morten Ravn (2010) for a recent discussion and references.

UNCERTAINTY SHOCK We use two external instruments. The first, motivated by Nicholas Bloom (2009), is the innovation in the VIX, where we use Bloom's (2009) series that links the VIX to other market uncertainty measures before the VIX was traded; the innovation is computed as the residual from an AR(2) (1962Q3–2011Q2).[17] The second is the innovation in the common component of the policy uncertainty index of Scott Baker, Bloom, and Steven Davis (2012), which is based on news

---

16. Other candidate instruments include the market announcement movements of Cochrane and Piazzesi (2002) and Faust, Swanson, and Wright (2004).

17. Lee, Rabanal, and Sandri (2010) take the uncertainty shock to be the innovation to the VIX.

media references to uncertainty in economic policy (1985Q1–2011Q2). The construction of measures of uncertainty is relatively new, and finding exogenous variation in uncertainty is challenging; for discussions see Geert Bekaert, Marie Hoerova, and Marco Lo Duca (2010) and Ruediger Bachman, Steffen Elstner, and Eric Sims (2010).

LIQUIDITY AND FINANCIAL RISK SHOCK   We use three external instruments. The first two are unadjusted and adjusted term spreads: the TED spread (1971Q1–2011Q2) and Gilchrist and Zakrajšek's (forthcoming) excess bond premium (1973Q3–2010Q3). Both instruments aim to measure risk in financial markets not associated with predictable default probabilities. The Gilchrist-Zakrajšek measure is a bond premium that has been adjusted to eliminate predictable default risk. For an early discussion of credit spreads as measures of market liquidity, see Benjamin Friedman and Kenneth Kuttner (1993); for more recent discussion see Gilchrist, Yankov, and Zakrajšek (2009) and Tobias Adrian, Paolo Colla, and Hyun Song Shin (forthcoming). The third instrument is William Bassett and coauthors' (2011) bank loan supply shock, which they compute as the unpredictable component of bank-level changes in lending standards, based on responses to the Federal Reserve's Senior Loan Officer Opinion Survey (1992Q1–2010Q4).

FISCAL POLICY SHOCK   We use three external instruments: Ramey's (2011a) federal spending news instrument (de-meaned, 1959Q1–2010Q4), Jonas Fisher and Ryan Peters's (2010) measure of excess returns on stocks of military contractors (1959Q1–2008Q4), and Romer and Romer's (2010) measure of tax changes relative to GDP ("all exogenous," de-meaned, 1959Q1–2007Q4). The first two of these are instruments for federal government spending changes, and the third is an instrument for federal tax changes. For additional discussion see Jonathan Parker (2011) and Ramey (2011b).

III.C. *Empirical Estimates of the Contribution of Various Shocks*

With these instruments in hand, we now undertake an empirical analysis of the contributions of the identified shocks to the 2007–09 recession. This analysis additionally requires a VAR for the factors, which we estimate using the "old" factors (see footnote 7). The VAR has four lags and is estimated over the full 1959Q1–2011Q2 sample.

HISTORICAL CONTRIBUTIONS AND CORRELATIONS   Table 6 summarizes the contributions to quarterly GDP growth of the 18 individually identified shocks (one shock per instrument) over the same subsamples as in table 1. Whereas the R^2s in table 1 measure the fraction of the variation in GDP

Table 6. Subsample R^2s of Regressions Evaluating the Factor Component of GDP Growth Associated with Individual Identified Shocks[a]

					R^2					
					Computed over 15 quarters starting at indicated cyclical peak					
Structural shock and instrument	F statistic[b]	1959–2007Q3	1984–2007Q3	1969Q4	1973Q4	1980Q1	1981Q3	1990Q3	2001Q1	2007Q4
Oil shock										
Hamilton	2.9	0.18	0.01	0.26	0.49	0.10	0.11	0.23	−0.55	−0.14
Kilian	1.1	0.08	−0.03	0.15	0.10	0.20	0.24	−0.03	−0.26	0.37
Ramey–Vine	1.8	0.14	−0.04	0.27	0.24	0.38	0.49	0.20	0.40	0.23
Monetary policy shock										
Romer–Romer	4.5	0.23	−0.16	0.35	0.31	0.57	0.57	0.16	0.08	0.26
Smets–Wouters	9.0	0.18	−0.01	0.24	0.25	0.32	0.23	0.37	0.16	0.46
Sims–Zha	6.5	0.19	−0.30	0.29	0.44	0.54	0.51	−0.01	0.03	0.03
Gürkaynak–Sack–Swanson	0.6	0.12	−0.07	0.07	0.39	0.26	0.26	0.00	−0.05	0.34
Productivity shock										
Fernald TFP	14.5	0.29	0.14	0.39	0.08	0.31	0.28	0.33	−0.42	−0.14
Galí long-run output per hour	NA	0.07	0.02	0.12	0.06	0.10	0.05	0.00	−0.11	−0.04
Smets–Wouters	7.0	0.20	−0.04	0.35	0.39	0.18	0.17	−0.05	−0.36	−0.17

(continued)

Table 6. Subsample R²s of Regressions Evaluating the Factor Component of GDP Growth Associated with Individual Identified Shocks[a] (*Continued*)

| | | R^2 | | | | | | | | | |
| | | | | Computed over 15 quarters starting at indicated cyclical peak | | | | | | | |
Structural shock and instrument	*F statistic[b]*	*1959–2007Q3*	*1984–2007Q3*	*1969Q4*	*1973Q4*	*1980Q1*	*1981Q3*	*1990Q3*	*2001Q1*	*2007Q4*
Uncertainty shock										
Bloom financial uncertainty (VIX)	43.2	0.08	0.03	0.12	0.16	0.18	0.31	0.22	0.23	0.34
Baker-Bloom-Davis policy uncertainty	12.5	0.11	0.02	0.15	0.40	0.15	0.14	0.48	0.15	0.62
Liquidity-financial risk shock										
Gilchrist-Zakrajšek spread	4.5	0.12	−0.09	0.17	0.21	0.25	0.38	0.07	0.30	0.57
TED spread	12.3	0.18	−0.08	0.28	0.45	0.31	0.21	0.21	0.30	0.38
Bassett and others bank loan supply	4.4	0.11	0.04	0.20	0.25	0.13	0.31	0.27	0.45	0.41
Fiscal policy shock										
Ramey spending	0.5	0.21	−0.06	0.36	0.35	0.50	0.55	0.06	0.01	0.19
Fisher-Peters spending	1.3	0.23	0.04	0.27	0.02	0.32	0.36	0.41	−0.06	0.09
Romer-Romer tax	0.5	0.16	−0.20	0.21	0.25	0.36	0.31	0.08	−0.38	−0.02
Principal components of uncertainty and credit spread shocks										
First principal component	NA	0.14	−0.03	0.21	0.36	0.25	0.26	0.30	0.23	0.48
Second principal component	NA	0.19	0.07	0.17	0.49	0.31	0.40	0.38	0.32	0.54

Source: Authors' calculations.

a. Structural shocks are computed using the full-sample six-factor DFM and the indicated instrument, as described in sections III.A and III.B of the text. R^2s are for the contribution of the indicated shock to the variation in GDP growth, computed over the indicated subsample, as described in footnote 19. NA = not applicable.

b. Non-HAC F statistic from the regression of the indicated instrument on the six factor innovations.

growth attributed to all the factors, the R^2s in table 6 measure the fraction of the variance attributed to current and past values of the individual row shock.[18] As in table 1, the R^2 is negative over subsamples in which the factor component arising from the identified shock covaries negatively with GDP growth. Additionally, the first column of table 6 reports the non-HAC (non-heteroskedasticity-and-autocorrelation-consistent) F statistic testing the hypothesis that the coefficients on the η_ts are all zero in the regression of Z_t on η_t, which is a measure of the strength of identification that enters the null distribution of the correlations in table 7 (Montiel Olea and others 2012).

As discussed in section III.A, the external instrument identification approach does not restrict the shocks to be uncorrelated. Table 7 reports the full-sample correlations among the shocks. If all the instruments within a category were identifying the same shock, and if the shocks were orthogonal, then the entries in the population version of table 7 would be 1 within categories and zero across categories. As discussed in section III.A, each shock is the predicted value from the regression of its instrument on the DFM innovations, so in general the correlation between two estimated shocks is different from the correlation between the instruments themselves.

Tables 6 and 7 suggest three main findings. First, for many of the external instruments, the F statistic in table 6, which is a measure of the strength of the instrument relevant to the distribution of the correlations in table 7, is small: it is less than 5 in 10 cases and less than 10 in all but 3. These small F statistics reinforce and extend Kilian's (2008b) observation that oil price shock series appear to be weak instruments. This suggests that there is considerable sampling uncertainty in the remaining statistics based on these instruments, but we do not attempt to quantify that uncertainty here.

Second, with this weak-instrument caveat, there is considerable variation of results across instruments within categories in tables 6 and 7. For example, whereas the correlation between the oil shocks identified using the Kilian (2008a) and Ramey-Vine (2010) instruments is 0.60, the correlation between the oil shocks identified using the Hamilton (2003) and Ramey-Vine (2010) instruments is only 0.15. Not surprisingly in light of these low correlations, the episode R^2s in table 6 vary considerably across

---

18. In the notation of equations 3 and 5, the factor component due to the jth structural shock is $\Lambda\Phi(L)^{-1}H_j\varepsilon_{jt}$. The R^2 of the ith variable with respect to the jth shock is thus computed as $R^2 = 1 - \left(\sum_t \left(\hat{e}_{it}^j\right)^2\right) \Big/ \left(\sum_t X_{it}^2\right)$, where $\hat{e}_{it}^j = X_{it} - \hat{\Lambda}_i' \hat{\Phi}(L)^{-1} \hat{H}_j\hat{\varepsilon}_{jt}$, where $\hat{\Lambda}_i'$ is the ith row of $\hat{\Lambda}$.

Table 7. Correlations among Estimated Structural Shocks

	O_H	O_K	O_{RV}	M_{RR}	M_{SW}	M_{SZ}	M_{GSS}	P_F	P_G	P_{SW}	U_B	U_{BBD}	L_{GZ}	L_{TED}	L_{BCDZ}	F_R	F_{FP}	F_{RR}
O_H	1.00																	
O_K	0.42	1.00																
O_{RV}	0.15	0.60	1.00															
M_{RR}	0.37	0.65	0.77	1.00														
M_{SW}	0.09	0.11	0.39	0.09	1.00													
M_{SZ}	0.33	0.35	0.68	0.93	0.16	1.00												
M_{GSS}	0.44	-0.12	-0.08	0.24	0.43	0.39	1.00											
P_F	-0.64	0.30	0.24	0.20	-0.09	0.06	-0.57	1.00										
P_G	-0.40	0.34	0.01	-0.30	0.35	-0.53	-0.37	0.52	1.00									
P_{SW}	-0.91	-0.03	0.00	-0.24	-0.07	-0.36	-0.59	0.82	0.68	1.00								
U_B	-0.37	-0.37	-0.58	-0.39	0.30	-0.29	0.37	0.19	0.34	0.27	1.00							
U_{BBD}	0.10	0.11	-0.37	-0.17	0.45	-0.22	0.57	-0.06	0.45	-0.01	0.78	1.00						
L_{GZ}	-0.20	-0.42	-0.51	-0.41	0.44	-0.24	0.34	0.07	0.24	0.08	0.92	0.66	1.00					
L_{TED}	-0.09	0.01	-0.05	0.03	0.73	0.10	0.48	0.21	0.37	0.09	0.80	0.76	0.84	1.00				
L_{BCDZ}	0.04	0.22	0.79	0.56	0.13	0.55	0.04	-0.09	-0.28	-0.06	-0.69	-0.54	-0.73	-0.40	1.00			
F_R	-0.17	-0.64	-0.77	-0.84	-0.32	-0.72	-0.34	-0.17	-0.01	0.01	0.26	-0.08	0.40	-0.13	-0.13	1.00		
F_{FP}	0.04	-0.21	-0.35	-0.72	0.20	-0.78	-0.03	0.04	0.40	-0.02	0.03	0.25	0.03	-0.12	-0.12	0.38	1.00	
F_{RR}	0.20	0.15	0.30	0.77	-0.10	0.88	0.37	0.18	-0.59	-0.28	0.01	-0.10	0.02	0.19	0.19	-0.45	-0.93	1.00

Source: Authors' calculations.

a. Entries are correlations between individually identified shocks, computed over the full 1959–2011Q2 sample. Shading denotes correlations within categories of shocks; italicized correlations are cross-category correlations exceeding 0.60 in absolute value. Shocks are those listed in table 6, abbreviated as follows:

Oil: O_H, Hamilton (1996, 2003); O_K, Kilian (2008a); O_{RV}, Ramey-Vine (2010)

Monetary policy: M_{RR}, Romer and Romer (2004); M_{SW}, Smets and Wouters (2007); M_{SZ}, Sims and Zha (2006); M_{GSS}, Gürkaynak, Sack, and Swanson (2005)

Productivity: P_F, Fernald (2009) TFP; P_G, Galí (1999); P_{SW}, Smets and Wouters (2007)

Uncertainty: U_{VIX}, Bloom (2009) financial uncertainty; U_{BBD}, Baker, Bloom, and Davis (2012) policy uncertainty

Liquidity and financial risk: L_{GZ}, Gilchrist and Zakrajšek (forthcoming) spread; L_{TED}, TED spread; L_{BCDZ}, Bassett and others (2011) bank loan supply

Fiscal policy: F_R, Ramey (2011a) spending; F_{FP}, Fisher-Peters (2010) spending; F_{RR}, Romer-Romer (2010) tax.

instruments within a shock category; for example, the Hamilton-identified oil shock has a subsample R^2 of -0.14 for 2007Q4–2011Q2, whereas for the Kilian-identified oil shock this R^2 is 0.37. Wide ranges of correlations are also evident among the four monetary policy shocks, although, interestingly, the variation in the subsample R^2s is less, with perhaps the exception of the Sims-Zha (2006) identified shock. Among fiscal policy shocks, the correlation between the shocks identified using the Ramey (2011a) and Romer and Romer (2010) instruments is -0.45 (the sign is negative because one is spending, the other tax), and the correlation between the Ramey (2011a) and Fisher-Peters (2010) spending shocks is only 0.38. In contrast, the correlation between the Fisher-Peters (2010) and Romer-Romer (2010) identified shocks is surprisingly large, -0.93, given that Fisher and Peters (2010) focus on exogenous changes in government spending whereas Romer and Romer (2010) focus on exogenous tax changes.[19]

The observation that the different instruments within a category identify different shocks with different effects echoes Glenn Rudebusch's (1998) critique of monetary policy shocks in structural VARs. One response is that these instruments are intended to estimate different effects; for example, the Romer-Romer fiscal instrument is intended to identify a tax shock, whereas the Ramey (2011a) and Fisher-Peters (2010) instruments are intended to identify spending shocks. Similarly, Kilian (2008b, 2009) argues that the Kilian (2008a) instrument estimates an oil supply shock, whereas the Hamilton instrument does not distinguish among the sources of price movements. Although the response that the different instruments are intended to estimate different shocks has merit, it then confronts the problem that the individually identified shocks within a category are not *un*correlated. For example, the correlation of -0.93 between the fiscal shocks identified by the Romer-Romer (2010) tax instrument and the Fisher-Peters (2010) spending instrument makes it problematic to treat these two shocks as distinct.

Third, again with the weak-instrument caveat, there is considerable correlation among individually identified shocks across categories of shocks, which suggests that superficially different instruments are capturing the same movements in the data (cross-category correlations exceeding 0.6 in absolute value are italicized in table 7). One notable set of correlations

19. Again, to be clear, these statements concern correlations among the shock series estimated using the instruments, not correlations among the underlying instruments themselves. For example, whereas the correlation between the shock estimated using the Fisher-Peters (2010) spending instrument and the shock estimated using the Romer-Romer (2010) tax instrument is -0.93, the correlation between the Fisher-Peters and the Romer-Romer instruments themselves is only -0.06.

is between the blocks of monetary and fiscal shocks: the mean average absolute correlation between individually identified shocks across the two categories is 0.51. The monetary and fiscal shock literatures speak of the difficulty of identifying one shock while holding the other constant, and this difficulty seems to arise in the large absolute correlations between the shocks from these two literatures.

Another notable block of large correlations is between the uncertainty shock and the liquidity and financial risk shock: the average absolute correlation between shocks across the two categories is 0.73. Indeed, for these categories the cross-category correlations are comparable to the within-category correlations. The subsample R^2s in table 6 also display similar patterns across these four identified shocks. It is perhaps not surprising that the VIX shock and the TED spread shock are correlated, because neither isolates a specific source for the shock; for example, a financial market disruption that both heightened uncertainty and increased financial sector risk would appear as shocks to both series. We find it more surprising that the correlation is 0.66 between the shocks identified using the Baker, Bloom, and Davis (2012) policy uncertainty index and the Gilchrist-Zakrajšek (forthcoming) excess bond premium spread. In any event, these two sets of instruments do not seem to be identifying distinct shocks. As a result, we also consider two composites of these five shocks constructed as the first two principal components of the five identified shocks. The subsample R^2s for the first principal component and for the first and second principal components combined are reported in the final rows of table 6.

THE 2007–09 RECESSION   Table 8 summarizes the contribution of the shocks in table 6 to the cumulative growth of GDP and of employment over three periods starting in 2007Q4. Because the shocks are correlated, these contributions are not an additive decomposition of the total factor component. Because all contributions and actuals are expressed in terms of deviations from trend, in 2011Q2 GDP remained 8.2 percent below its trend value, extrapolated from the 2007Q4 peak, of which 6.0 percentage points was the contribution of the factors. Plots of the contributions of the individual shocks over the full sample, along with the shock contributions to other variables, are presented in the online appendix.

Consider the recession period, 2007Q4–2009Q2. The largest negative shock contributions to the drops in GDP and employment are seen in the financial shock measures (the liquidity-risk and uncertainty shocks). The composite uncertainty-liquidity shock based on the first principal component of the five estimated shocks in this category attributes approximately two-thirds of the recession's decline in GDP and employment to financial

Table 8. Contributions of Identified Shocks to Cumulative Post-2007Q4 Growth of GDP and Employment

Shock and instrument	Cumulative percentage change in detrended GDP[a]			Cumulative percentage change in detrended payroll employment[a]				2009Q2 forecast growth in factor component, 2009Q2–2011Q2[b]	
	2007Q4–2008Q3	2007Q4–2009Q2	2007Q4–2011Q2	2007Q4–2008Q3	2007Q4–2009Q2	2007Q4–2011Q2		GDP	Employment
Actual outcome	−2.8	−8.7	−8.2	−1.4	−6.2	−7.4			
Factor component	−4.0	−9.2	−6.0	−2.1	−7.3	−8.9			
Oil shock									
Hamilton	−1.0	−0.8	−0.8	−0.4	−1.1	−1.0		1.8	1.0
Kilian	−1.0	−2.1	−0.6	−0.5	−1.7	−1.2		1.3	0.3
Ramey–Vine	−1.4	−1.6	0.7	−1.1	−2.1	0.2		1.5	1.2
Monetary policy shock									
Romer–Romer	−1.1	−1.6	0.1	−0.7	−2.1	0.4		1.4	1.2
Smets–Wouters	−0.4	−3.7	−5.0	−0.5	−2.5	−6.2		−0.6	−3.3
Sims–Zha	−0.3	−0.1	−0.1	−0.5	−1.0	0.3		0.4	0.8
GSS	−0.8	−3.2	−4.3	0.0	−1.0	−3.8		−1.3	−2.7
Productivity shock									
Fernald	−0.7	0.0	1.4	−0.1	0.1	0.9		0.2	0.2
Galí	0.4	0.3	1.1	0.0	−0.6	−0.9		1.0	0.5
Smets–Wouters	−0.5	−0.1	0.6	0.2	−0.1	0.4		0.9	0.7

(continued)

Table 8. Contributions of Identified Shocks to Cumulative Post-2007Q4 Growth of GDP and Employment (*Continued*)

Shock and instrument	Cumulative percentage change in detrended GDP[a]			Cumulative percentage change in detrended payroll employment[a]			2009Q2 forecast growth in factor component, 2009Q2–2011Q2[b]	
	2007Q4–2008Q3	2007Q4–2009Q2	2007Q4–2011Q2	2007Q4–2008Q3	2007Q4–2009Q2	2007Q4–2011Q2	GDP	Employment
Uncertainty shock								
Financial uncertainty (VIX)	−1.0	−4.1	0.2	−0.9	−3.7	−2.2	3.5	0.6
Political uncertainty (BBD)	−2.3	−5.8	−4.8	−1.6	−4.6	−6.8	1.4	−2.0
Liquidity/financial risk shock								
Gilchrist-Zakrajšek spread	−1.5	−6.3	−1.2	−0.8	−4.6	−3.5	3.6	−0.4
TED spread	−1.2	−5.6	−4.9	−0.8	−3.7	−6.8	0.8	−3.2
Bassett and others bank loan	−2.0	−3.2	0.1	−1.5	−3.5	−1.1	2.7	1.3
Fiscal policy shock								
Ramey spending	−1.6	−1.6	−1.0	−0.9	−1.6	−0.5	0.3	0.1
Fisher-Peters spending	−0.3	−0.3	0.3	−0.2	−0.7	0.3	0.4	0.5
Romer-Romer tax	0.2	0.3	−0.2	0.2	0.0	0.3	−0.1	0.3
Principal components of uncertainty and credit spread shocks								
First principal component	−1.5	−6.2	−3.4	−1.0	−4.5	−5.8	2.4	−1.8
Second principal component	−2.3	−7.6	−5.4	−1.5	−5.8	−8.4	2.7	−2.5

Source: Authors' calculations.

a. Each entry after the top panel is the contribution of the indicated shock to cumulative growth in GDP or employment over the indicated period, computed as described in footnote 19 in the text.

b. Implied forecasts of GDP or employment growth from 2009Q2 to 2011Q2, constructed in 2009Q2 associated with the indicated shock.

factors (6.2 of 9.2 percentage points and 4.5 of 7.3 percentage points, respectively). Oil shocks and monetary policy shocks both make moderate negative contributions, with the exception of the Sims-Zha (2006) identified shock. The Romer-Romer (2004), Smets-Wouters (2007), and Gürkaynak, Sack, and Swanson (2005) shocks indicate that monetary policy was neutral or contractionary during the recession and recovery, which is consistent with the model being linear and not incorporating a zero lower bound (so the federal funds rate, which cannot drop below zero, exerted a contractionary effect). Unfortunately, our identification scheme does not capture the unconventional monetary policy of the crisis and recovery. During 2007Q4–2009Q2, the effects of productivity and fiscal policy shocks on GDP growth are estimated to be small.

III.D. Discussion

Inference about the causes of the 2007–09 recession based on table 8 is complicated because the different instruments identify shocks that in several cases have a low correlation within category, and in other cases have high correlations across categories. Because our approach is to adopt identification schemes from the literature, this suggests internal inconsistencies in the identified VAR literature concerning individual identified shocks. Perhaps to oversimplify, what some authors call a monetary policy shock looks much like what other authors call a fiscal policy shock, and what some authors call an uncertainty shock looks much like what others call a liquidity or excess financial risk shock. These puzzling results might come about because our analysis is insufficiently nuanced to distinguish between the different estimands of the different instruments, or because we have too few factors to span the space of the potentially many structural shocks, or because of large sampling uncertainty arising from weak instruments. In any event, the low correlations among some of the monetary policy shocks, the high correlations between the monetary and fiscal policy shocks, and the high correlations among the uncertainty and the liquidity/financial risk shocks preclude a compelling decomposition.

Despite this substantial caveat, some substantive results emerge from tables 6 and 8. The contributions of productivity, monetary policy, and fiscal policy shocks to the 2007–09 recession are small. Oil shocks contributed to the decline, especially before the financial crisis. The main contributions to the decline in output and employment during the recession are estimated to come from financial and uncertainty shocks. The plot of the contribution of the first principal component of these five individually identified shocks (figure 3) shows that they explain a great deal of

Figure 3. Contribution to Detrended Four-Quarter Growth of GDP and Nonfarm Employment of the First Principal Component of the Identified Financial and Uncertainty Shocks[a]

GDP

Percent

Actual

Principal component

Nonfarm business employment
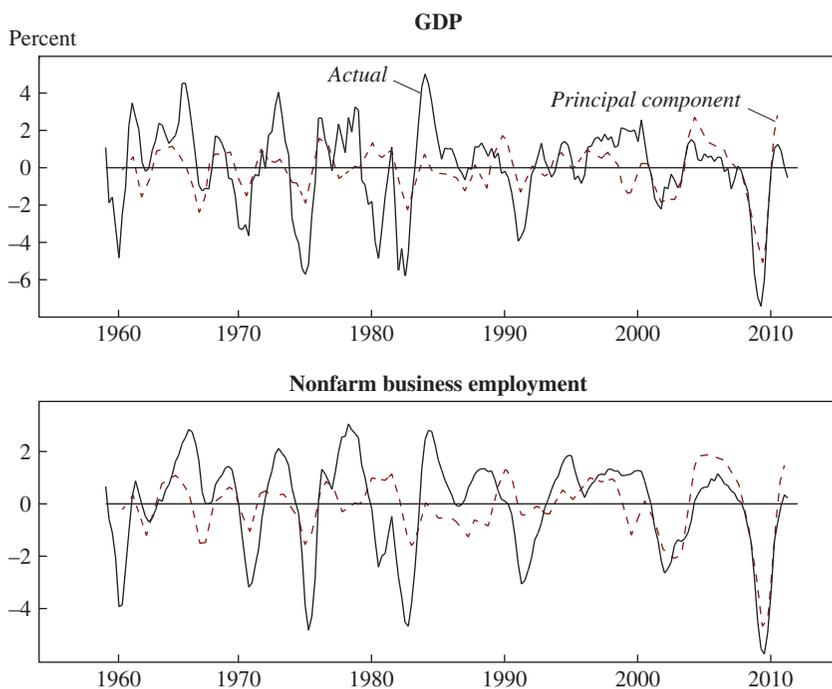
Source: Authors' calculations.
a. The five identified shocks are the three liquidity and financial risk shocks and the two uncertainty shocks listed in table 6.

the 2007–09 recession and subsequent recovery, and that they also play an important but lesser role in prior fluctuations. Taken at face value, this suggests an economy being hit by a sequence of unusually large shocks, all of which have been experienced before, but not with such magnitude or in such close succession: oil shocks initially, followed by the financial crisis, financial market disruptions, and a prolonged period of uncertainty.

IV. The Slow Recovery

On its face, the unusually slow recovery following the 2009Q2 trough seems inconsistent with the conclusion of the previous sections that the macroeconomic dynamics of this recession are consistent with those of

prior recessions, simply with larger shocks. Indeed, during the 8 quarters following that trough, GDP grew by only 5.0 percent, compared with an average of 9.2 percent after 8 quarters for the recessions from 1960 through 2001, and employment increased by only 0.6 percent, compared with a 1960–2001 average of 4.0 percent.[20] The contrast between the current slow recovery and the robust recoveries of 1960–82 is even more striking: those recessions averaged 8-quarter GDP growth of 11.0 percent and 8-quarter employment growth of 5.9 percent following the trough. In this section we therefore take a closer look at the extent to which the current slow recovery is or is not consistent with historical experience.

Why has the recovery in employment since 2009Q2 been so much slower than the 1960–82 recoveries? In the context of the DFM, employment growth after a trough is the sum of four terms: trend employment growth, the predicted cyclical common component (deviations from trend) given the state of the economy at the trough, the prediction errors in the cyclical common component, and the series-specific idiosyncratic errors. Accordingly, the weakness of the recovery since 2009Q2 relative to, say, that after 1982Q4 could arise from differences in underlying trends (such as in demographics), differences in recovery paths after different types of recessions (such as one induced by monetary policy versus one following a financial crisis), differences in macroeconomic luck once the recovery commenced, or peculiarities of employment unrelated to the rest of the economy. Although the latter two terms might be of historical interest, the former two shed more light on structural differences between the two recoveries. In this section we therefore focus on the first two of these terms—the trend and the predicted cyclical common component—comparing their values in the post-2009Q2 recovery with their values in previous recoveries.

As in section III, the calculations here require a VAR for the factors, which we estimate using four lags and the "old" factors over the 1959–2007Q3 period. With this model held constant, differences in the predicted cyclical component across recoveries reflect differences in recovery paths implied by the shocks that produced the recessions. This permits a decomposition of the slow pace of the recovery after 2009Q2, relative to previous recoveries, into changes in the trend plus changes in the predicted cyclical component.

---

20. These averages exclude the recovery that began in 1980Q3 because the next recession started within the 8-quarter window of these calculations.

IV.A. *Different Shocks Imply Different Recovery Paths*

Different structural shocks induce different macroeconomic responses. For example, Bloom (2009) predicts a fast recovery after an uncertainty shock (investment and consumption pick up as soon as the uncertainty is resolved), whereas Carmen Reinhart and Kenneth Rogoff (2009) describe recoveries from financial crises as typically slow. In terms of the factor model, the state of the economy at the trough is summarized by the current and past values of the factors as of the trough. Because the values of the shocks (and thus of the factors) vary across recessions both in composition and in magnitude, the recovery paths predicted by the DFM also vary across recessions.

Figure 4 plots actual quarterly employment growth, its common component, and its predicted common component following each of the eight post-1960 troughs. All series are expressed as deviations from trend, so that a value of zero denotes employment growth at trend. The predicted common component is computed using the values of the factors through the trough date; that is, the predicted common component is the forecast of the common component one would make standing at the trough, given the historical values of the factors through the trough date and the model parameters (because the DFM was estimated through 2007Q3, the predicted components in figure 4 are in-sample for the first seven recessions and pseudo-out-of-sample forecasts for 2009Q2). The difference between actual employment growth and its common component is the idiosyncratic disturbance (e_t in equation 1). The difference between the common component and the predicted common component arises from the factor innovations (η_t in equation 2) that occurred after the trough.

Three features of figure 4 are noteworthy. First, there is considerable heterogeneity across recessions in both the shape and the magnitude of predicted recoveries of employment. By construction, the sole source of this heterogeneity is differences in the state of the economy, as measured by the factors, at the trough. Strong positive employment growth is predicted following the 1982Q4 trough—employment growth returns to trend only 3 quarters after the trough—whereas slow employment growth is predicted following 1980Q3, 1991Q1, and 2009Q2.

Second, in most recessions the predicted values track the actual common component. The main exception is the 1980Q3 recovery, which was interrupted early on by the next recession.

Third, given the values of the factors in 2009Q2, the DFM predicts 6 quarters of subtrend employment growth following the 2009Q2 trough. In

Figure 4. Deviations of Employment Growth from Trend Following Cyclical Troughs, 1961–2011[a]
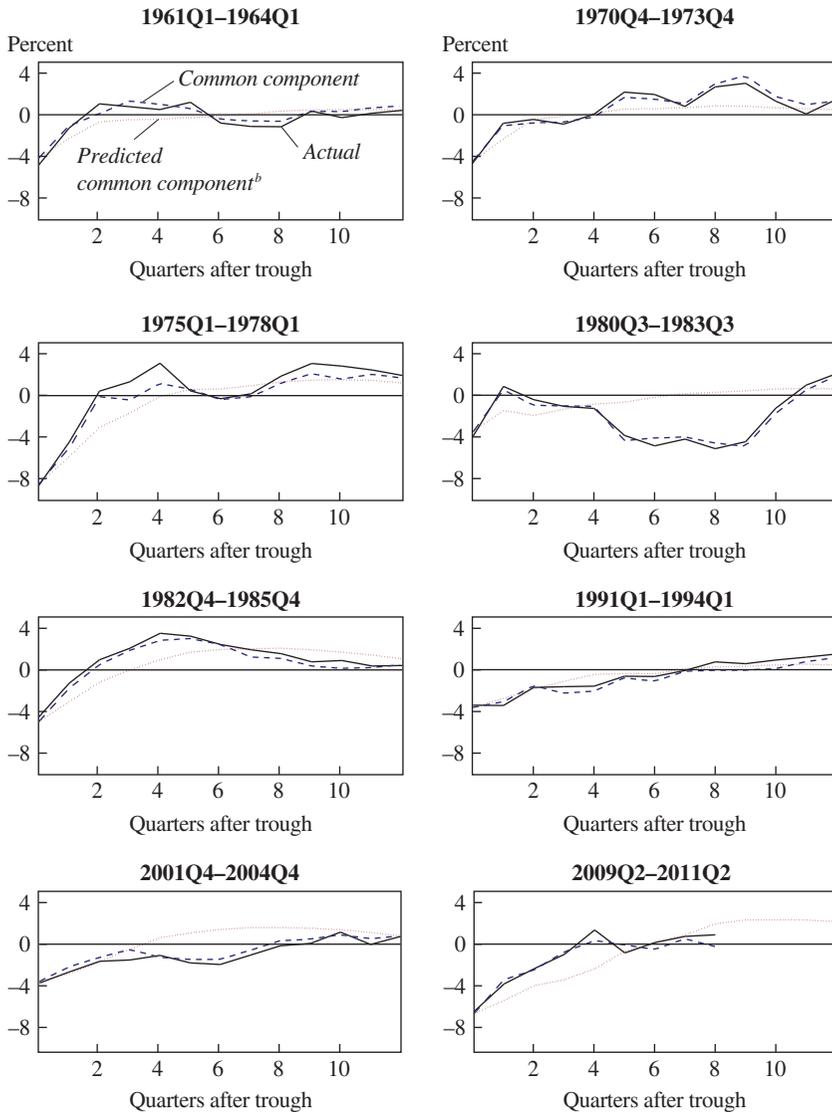
Source: Authors' calculations.
a. Growth rates are annualized quarterly rates. A value of zero indicates no deviation from trend.
b. Predicted based on the factors at the trough.

fact, the DFM predicts a slower employment recovery from the 2009Q2 trough than actually occurred; that is, the current recovery in employment is actually *stronger* than predicted.[21]

IV.B. Decomposition of the Post-2009Q2 Recovery into Trend and Cyclical Components

We now turn to the decomposition of the post-1959 recoveries into their trend and predicted cyclical components, where the latter are computed as described in the previous subsection using the factors at the trough. Table 9 summarizes the results for 8-quarter cumulative post-trough growth of GDP, employment, and productivity.

Consistent with the trends plotted in figure 1, table 9 shows that the trend component of predicted growth in GDP and employment falls over time. Consistent with the cyclical components plotted in figure 4, there is considerable variation in the predicted cyclical components, which arises from variation in the composition and magnitude of the factors at the trough. The predicted cyclical contributions to 8-quarter employment growth range from +1.1 percentage points following the 1982Q4 trough to −3.1 percentage points following the 2009Q2 trough.

The final rows of table 9 report the decomposition into trend and cycle of the difference between the predicted 8-quarter growth following 2009Q2 and the corresponding averages for pre-1984 recoveries. Predicted GDP growth emerging from 2009Q2 is 3.0 percentage points less than the pre-1984 average; four-fifths of this gap (2.4 percentage points) is due to differences in trend. Predicted employment growth is 6.0 percentage points less than the pre-1984 average; of this gap, 2.7 percentage points is attributed to differences in the cyclical components, whereas most, 3.3 percentage points, is attributed to differences in trend employment growth. The predicted cyclical component of productivity growth in the post-2009Q2 recovery is unusually large, 6.3 percentage points, although this predicted value is perhaps comparable to its values in the recoveries after 1975Q1 and 1982Q4. The difference between the trend components of productivity growth in the recovery after 2009Q2 and in the average for the pre-1984 recoveries is 0.5 percentage point; that is, trend productivity growth in the post-2009Q2 episode is slightly higher than its 1960–82 average. Most of the difference in productivity growth

---

21. Allowing for a break in $\Phi(L)$ in 1984Q1 produces somewhat faster predicted recoveries before 1984 and somewhat slower ones after 1984; for details see the online appendix.

Table 9. Predicted and Actual Cumulative Growth of Output, Employment, and Productivity Following a Cyclical Trough[a]

Percent

Trough	Source	GDP	Nonfarm employment	Output per hour (nonfarm business)
			Cumulative growth of common component (or actual) over 8 quarters following trough	
1961Q1	Cyclical	1.1	−1.0	2.0
	Trend	7.5	4.9	4.8
	Total	8.7	4.0	6.8
1970Q4	Cyclical	2.4	0.0	2.6
	Trend	6.9	4.7	4.0
	Total	9.3	4.6	6.6
1975Q1	Cyclical	3.3	−1.8	5.4
	Trend	6.6	4.5	3.7
	Total	9.9	2.7	9.1
1980Q3	Cyclical	1.1	−1.5	2.9
	Trend	6.3	4.2	3.5
	Total	7.5	2.7	6.4
1982Q4	Cyclical	5.0	1.1	4.3
	Trend	6.2	4.1	3.5
	Total	11.2	5.2	7.8
1991Q1	Cyclical	0.8	−1.6	2.5
	Trend	5.9	3.3	3.8
	Total	6.7	1.6	6.3
2001Q4	Cyclical	2.9	0.5	2.6
	Trend	5.1	2.1	4.3
	Total	8.0	2.6	6.9
2009Q2	Cyclical	2.4	−3.1	6.3
	Trend	4.4	1.2	4.5
	Total	6.8	−1.9	10.8
Averages				
1960–82	Cyclical	3.0	−0.4	3.6
	Trend	6.8	4.5	4.0
	Total	9.8	4.1	7.6
	Actual[b]	11.0	5.9	7.3
1960–2001	Cyclical	2.6	−0.5	3.2
	Trend	6.4	3.9	4.0
	Total	9.0	3.5	7.3
	Actual[b]	9.2	4.0	7.2
Differences				
2009Q2 minus	Cyclical	−0.6	−2.7	2.7
average, 1960–82	Trend	−2.4	−3.3	0.5
	Total	−3.0	−6.0	3.2

Source: Authors' calculations.

a. Entries are cumulative predicted growth rates of the common component (or the actual value) of the indicated series, computed using the factors at the trough and the DFM estimated through 2007Q3. Predicted paths are decomposed into the detrended cyclical component (the contribution of the factors at the trough) and the trend growth rate.

b. Excludes the post-1980Q3 recovery because the next recession commenced within the 8-quarter window used in this table.

between the post-2009Q2 recovery and the 1960–82 recoveries is attributed to differences in the cyclical component.[22]

IV.C. The Slowdown in Trend Labor Force Growth and Slow Recoveries

A striking result of the previous section is that the decline in the trend component accounts for nearly all of the slowdown in GDP growth, and for over half the slowdown in employment growth, in the current recovery relative to the pre-1984 averages. Table 10 decomposes the change in trend GDP growth from 1965 to 2005 into GDP per employee, the employment-population ratio, the labor force participation rate, and the growth of the labor force. As seen in the first panel of table 10, the decline in the trend growth rate of GDP of 1.2 percentage points from 1965 to 2005 is, in this accounting sense, almost entirely due to declines in trend employment, which in turn is approximately equally due to declines in growth of the employment-population ratio and in population growth. In this accounting sense, the third panel of the table shows that declines in the growth of the employment-population ratio are in turn due to declines in the growth of the labor force participation rate, which in turn are largely due to declines in the growth rate of the female labor force participation rate. Figure 5 presents the estimated trends for the terms in the first panel in table 10 for the full 1959–2011 period.

Because the trend value of the unemployment rate is approximately the same in the 1960s as in the early 2000s (after peaking in the early 1980s), understanding the decline in mean employment growth amounts to understanding the decline in the growth of the labor force.[23] A significant literature examines long-term labor force trends and links them to two major

---

22. The online appendix reports results for five- and seven-factor models. The only notable departure from the results reported in this paper for the six-factor model is that the five- and seven-factor models predict a stronger post-2009Q2 recovery, so they attribute even more of the gap between that recovery and the 1960–82 recoveries to the slowdown in trend growth.

23. Two pieces of evidence suggest that the observed decline in employment growth is not an artifact of long-term mismeasurement. First, trend growth in employment measured by the household survey exhibits the same pattern as that in the establishment survey, with a decline from 2.1 percent annually in 1970 to 1.0 percent annually in 2000; this 1.1-percentage-point decline is close to the 1.4-percentage-point decline in the establishment survey (see the online appendix). Second, the small net trend in GDP per worker (from the establishment survey) matches the small net trend in output per hour (nonfarm business), which would not be the case if nonfarm business hours (a narrower measure) are correctly measured but employment is increasingly underestimated.

Table 10. Contributions of Trend Productivity, Labor Force, and Population to Trend GDP Growth Rate[a]

Series and component	Trend growth rate[b] (percent per year)			Difference, 2005 minus 1965[c] (percentage points)
	1965	*1985*	*2005*	
GDP	3.7	3.1	2.5	−1.2
GDP-employment ratio	1.6	1.3	1.5	−0.1
Employment-population ratio	0.3	0.4	−0.2	−0.5
Population	1.7	1.4	1.1	−0.6
GDP-employment ratio	1.6	1.3	1.5	−0.1
Ratio of GDP to NFB output	−0.2	−0.3	−0.2	0.0
Ratio of NFB output to NFB hours	2.3	1.8	2.2	−0.1
Ratio of NFB hours to NFB employment	−0.4	−0.3	−0.2	0.2
Ratio of NFB employment to total nonfarm employment	0.0	0.1	−0.3	−0.3
Employment-population ratio	0.3	0.4	−0.2	−0.5
Employment as share of labor force	0.0	0.1	0.0	0.0
Labor force as share of population	0.3	0.4	−0.1	−0.5
Labor force share of population	0.3	0.4	−0.1	−0.5
Female	0.5	0.4	0.0	−0.5
Male	−0.2	−0.1	−0.2	0.1
Labor force	2.0	1.7	0.9	−1.1
Female (prime-age)	0.7	0.8	0.3	−0.4
Male (prime-age)	0.4	0.6	0.2	−0.2
Female (non-prime-age)	0.5	0.2	0.2	−0.3
Male (non-prime-age)	0.4	0.1	0.2	−0.2

Source: Authors' calculations.

a. Growth rates and differences for components may not sum to those for totals because of rounding. Standard errors for the estimated trends range from 0.1 for the labor force variables to 0.5 for GDP; for details see the online appendix. NFB = nonfarm business.

b. Each entry is the growth in the trend component of the indicated series in the indicated year, computed as described in section I.C.

c. Difference between 2005 and 1965 trend values.

Figure 5. Trend Components of Growth of GDP, the GDP-Employment Ratio, the Employment-Population Ratio, and Population, 1959–2011
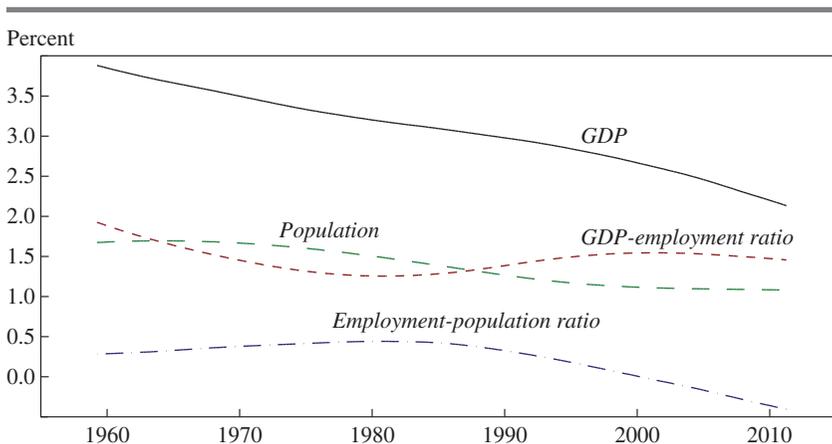
Source: Authors' calculations.

Figure 6. Civilian Labor Force Participation Rates, Total and by Sex, 1959–2011

Percent
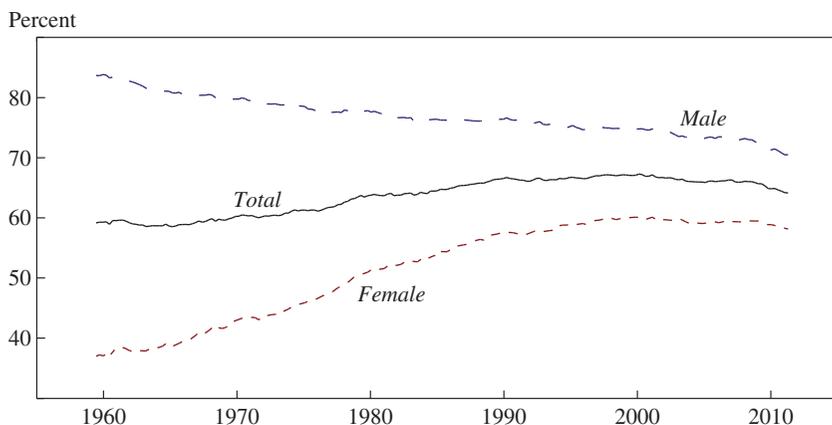
Source: Bureau of Labor Statistics.

demographic shifts (figure 6).[24] The first is the historic increase in the female labor force participation rate from the 1960s through the 1990s and its subsequent plateau; see Claudia Goldin (2006) for an extensive discussion. The second is the (smaller) decline in the male labor force participation rate. Stephanie Aaronson and others (2006) and Bruce Fallick and Jonathan Pingle (2008) attribute this decline to a combination of changes in the age distribution of workers and changing cohort labor force participation rates associated with the aging of the baby-boom generation (also see Fallick, Fleischman, and Pingle 2010). The main conclusion from this demographic work is that, barring a new surge in female labor force participation or a significant increase in the growth rate of the population, these demographic factors point toward a further decline in trend growth of

24. Focusing solely on demographic shifts ignores other potential factors affecting labor force participation. One such factor is an endogenous response to the stagnation of median real wages; however, although the magnitude of the labor supply elasticity is debated, micro studies generally suggest that it is small (see Saez, Slemrod, and Giertz 2012, Chetty 2011, and Chetty and others forthcoming for discussions). Another such factor is a possible trend increase in the mismatch between worker skills and available jobs. For example, Goldin and Katz (2008) point to a plateau in the supply of educated Americans around 1980. Jaimovich and Siu (2012) present evidence that the trend adjustments in employment occur mainly through permanent losses of mid-skill jobs during and following recessions; this view of step-like adjustments differs from our smooth trend. It goes beyond the scope of this paper to examine these factors in any detail.

employment and hours in the coming decades. Applying this demographic view to recessions and recoveries suggests that future recessions with historically typical cyclical behavior will have steeper declines and slower recoveries in output and employment.

V. Conclusions and Discussion

Three main substantive conclusions emerge from this work. First, the recession of 2007–09 was the result of shocks that were larger versions of shocks previously experienced, to which the economy responded in a historically predictable way. Second, these shocks emanated primarily, but not exclusively, from financial upheaval and heightened uncertainty. Third, although the slow nature of the subsequent recovery is partly due to the nature and magnitude of the shocks that caused the recession, most of the slow recovery in employment, and nearly all of that in output, is due to a secular slowdown in trend labor force growth. This slowdown provides a simple explanation for the jobless recoveries of the 2001 and 2007–09 recessions. To the extent that it derives (as the literature suggests) from persistent demographic changes, recoveries from future recessions can be expected to be "jobless" as well. To these substantive conclusions we would add a fourth, methodological conclusion: that ignoring these changing trends will impart low-frequency movements to the errors, which seems likely to introduce subtle problems into structural VAR analysis.

The above three substantive conclusions are subject to a number of caveats. First, although the evidence for the stability of the factor loadings is relatively strong, it is difficult to draw inferences about the stability of the factor VAR parameters with only 15 quarters of post-2007Q3 data, particularly in the presence of evident heteroskedasticity in the factor innovations. The fact that the current recovery in employment has been stronger than predicted by the DFM given the state of the economy at the trough could reflect the effectiveness of the extraordinary monetary and fiscal policy measures taken during the recession, or it could be an indication of parameter instability; we are unable to distinguish between these two possibilities with the current limited data.

Second, the structural DFM analysis using the method of external instruments estimates shocks that are correlated with each other. The ability to estimate this correlation, rather than needing to impose it as an identifying restriction, is a strength of this methodology. Finding sometimes-large correlations across different types of shocks suggests that different identification strategies are estimating similar features of

the data, but interpreting them differently. This raises broader challenges for the structural DFM and VAR literatures, which lie beyond the scope of this analysis. Sorting out credible instrumental variables methods for separately identifying liquidity shocks, market risk shocks, exogenous wealth shocks, and uncertainty shocks constitutes a large research agenda.

References

Aaronson, Stephanie, Bruce Fallick, Andrew Figura, Jonathan Pingle, and William Wascher. 2006. "The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply." *BPEA*, no. 1: 69–154 (with discussion).

Adrian, Tobias, Paolo Colla, and Hyun Song Shin. Forthcoming. "Which Financial Frictions? Parsing the Evidence from the Financial Crisis of 2007–9." *NBER Macroeconomics Annual.*

Andrews, Donald W. K. 2003. "End-of-Sample Instability Tests." *Econometrica* 71, no. 6: 1661–94.

Bachmann, Ruediger, Steffen Elstner, and Eric Sims. 2010. "Uncertainty and Economic Activity: Evidence from Business Survey Data." Working Paper no. 16143. Cambridge, Mass.: National Bureau of Economic Research.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70: 191–221.

———. 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica* 74:1133–50.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2012. "Measuring Economic Policy Uncertainty." Stanford University.

Bank of Canada. 2011. *Monetary Policy Report: October 2011.* Ottawa.

Bassett, William F., Mary Beth Chosak, John C. Driscoll, and Egon Zakrajšek. 2011. "Changes in Bank Lending Standards and the Macroeconomy." Washington: Division of Monetary Affairs, Board of Governors of the Federal Reserve System.

Basu, Susanto, John Fernald, and Miles Kimball. 2006. "Are Technology Improvements Contractionary?" *American Economic Review* 96, no. 5: 1418–48.

Bates, Brandon J., Mikkel Plagborg-Møller, James H. Stock, and Mark W. Watson. 2012. "Consistent Factor Estimation in Dynamic Factor Models with Structural Instability." Harvard University.

Baxter, Marianne B., and Robert G. King. 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *Review of Economics and Statistics* 81, no. 4: 575–93.

Bekaert, Geert, Marie Hoerova, and Marco Lo Duca. 2010. "Risk, Uncertainty, and Monetary Policy." Working Paper no. 16397. Cambridge, Mass.: National Bureau of Economic Research.

Berger, David. 2011. "Countercyclical Restructuring and Jobless Recoveries." Yale University.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120: 387–422.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77: 623–85.

Boivin, Jean, and Marc Giannoni. 2010. "DSGE Models in a Data-Rich Environment." Working Paper no. 12772. Cambridge, Mass.: National Bureau of Economic Research.

Bordo, Michael D., and Joseph G. Haubrich. 2011. "Deep Recessions, Fast Recoveries, and Financial Crises: Evidence from the American Record." Rutgers University.

Campbell, John Y., Stefano Giglio, and Christopher Polk. 2010. "Hard Times." Working Paper no. 16222. Cambridge, Mass.: National Bureau of Economic Research.

Chetty, Raj. Forthcoming. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica.*

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. Forthcoming. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." *NBER Macroeconomics Annual.*

Cochrane, John H., and Monica Piazzesi. 2002. "The Fed and Interest Rates: A High-Frequency Identification." *American Economic Review* 92: 90–95.

Eickmeier, Sandra, and Christina Ziegler. 2008. "How Successful Are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach." *Journal of Forecasting* 27, no. 3: 237–65.

Fallick, Bruce, and Jonathan Pingle. 2008. "The Effect of Population Aging on Aggregate Labor Supply in the United States." In *The Outlook for Labor Supply in the United States.* Federal Reserve Bank of Boston (with discussion).

Fallick, Bruce, Charles Fleischman, and Jonathan Pingle. 2010. "The Effect of Population Aging on the Aggregate Labor Market." In *Labor in the New Economy,* edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press for the National Bureau of Economic Research.

Faust, Jon, Eric Swanson, and Jonathan Wright. 2004. "Identifying VARs Based on High-Frequency Futures Data." *Journal of Monetary Economics* 51, no. 6: 1107–31.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *The Economic Journal* 120: 414–36.

Friedman, Benjamin, and Kenneth Kuttner. 1993. "Why Does the Paper-Bill Spread Predict Real Economic Activity?" In *Business Cycles, Indicators and Forecasting,* edited by J. H. Stock and M. W. Watson. University of Chicago Press for the National Bureau of Economic Research.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89: 249–71.

Giannone, Domenico, Michele Lenza, and Lucrezia Reichlin. 2012. "Money, Credit, Monetary Policy and the Business Cycle in the Euro Area." ECARES Working Paper no. 2012–008. Brussels: European Centre for Advanced Research in Economics and Statistics.

Giannone, Domenico, Lucrezia Reichlin, and Luca Sala. 2004. "Monetary Policy in Real Time." *NBER Macroeconomics Annual* 161–200.

Gilchrist, Simon, and Egon Zakrajšek. Forthcoming. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review.*

Gilchrist, Simon, V. Yankov, and Egon Zakrajšek. 2009. "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets." *Journal of Monetary Economics* 56: 471–93.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96, no. 2: 1–21.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Harvard University Press.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1, no. 1: 55–93.

Hall, Robert E. 2010. "Why Does the Economy Fall to Pieces after a Financial Crisis?" *Journal of Economic Perspectives* 24: 3–20.

———. 2011. "The Long Slump." *American Economic Review* 101: 431–69.

———. 2012. "Quantifying the Forces Leading to the Collapse of GDP after the Financial Crisis." Stanford University.

Hamilton, James D. 1996. "This Is What Happened to the Oil Price-Macroeconomy Relationship." *Journal of Monetary Economics* 38, no. 2: 215–20.

———. 2003. "What Is an Oil Shock?" *Journal of Econometrics* 113: 363–98.

———. 2009. "Causes and Consequences of the Oil Shock of 2007–08." *BPEA,* no. 1: 215–83.

———. 2010. "Nonlinearities and the Macroeconomic Effects of Oil Prices." Working Paper no. 16186. Cambridge, Mass.: National Bureau of Economic Research.

Jordà, Òscar, Moritz H. P. Schularick, and Alan M. Taylor. 2011. "When Credit Bites Back: Leverage, Business Cycles, and Crises." Working Paper no. 17621. Cambridge, Mass.: National Bureau of Economic Research.

Kilian, Lutz. 2008a. "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, no. 2: 216–40.

———. 2008b. "The Economic Effect of Energy Price Shocks." *Journal of Economic Literature* 46, no. 4: 871–909.

———. 2009. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review* 99, no. 3: 1053–69.

Kim, Chang-Jin, and Yungjo Eo. 2012. "Markov-Switching Models with Evolving Regime-Specific Parameters: Are Post-War Booms and Recessions All Alike?" University of Washington.

King, Robert G., and Mark W. Watson. 2012. "Inflation and Unit Labor Costs." Princeton University.

Lee, Jaewoo, Paul Rabanal, and Damiano Sandri. 2010. "U.S. Consumption after the 2008 Crisis." IMF Staff Position Note no. SPN/10/01. Washington: International Monetary Fund.

Lettau, Martin, and Sydney C. Ludvigson. 2011. "Shocks and Crashes." Working Paper no. 16996. Cambridge, Mass.: National Bureau of Economic Research.

Mertens, Karel, and Morten O. Ravn. 2010. "Technology-Hours Redux: Tax Changes and the Measurement of Technology Shocks." Cornell University.

———. 2012. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." Cornell University.

Mian, Atif, and Amir Sufi. 2011. "What Explains High Unemployment? The Aggregate Demand Channel." University of Chicago Booth School of Business.

Mishkin, Frederic S. 2010. "Over the Cliff: From the Subprime to the Global Financial Crisis." Working Paper no. 16609. Cambridge, Mass.: National Bureau of Economic Research.

Montiel Olea, José Luis, James H. Stock, and Mark W. Watson. 2012. "Inference in Structural VARs with External Instruments." Harvard University.

Parker, Jonathan. 2011. "On Measuring the Effects of Fiscal Policy in Recessions." *Journal of Economic Literature* 49, no. 3: 703–18.

Ramey, Valerie A. 2011a. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126, no. 1: 1–50.

———. 2011b. "Can Government Purchases Stimulate the Economy." *Journal of Economic Literature* 49, no. 3: 673–85.

Ramey, Valerie A., and Daniel J. Vine. 2010. "Oil, Automobiles, and the U.S. Economy: How Much Have Things Really Changed?" *NBER Macroeconomics Annual* 25: 333–67.

Reinhart, Carmen M., and Vincent R. Reinhart. 2010. "After the Fall." In *Macroeconomic Policy: Post-Crisis and Risks Ahead, Proceedings of the Federal Reserve Bank of Kansas City 2010 Jackson Hole Symposium.*

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Romer, Christina D., and David H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual* 4: 121–70.

———. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94: 1055–84.

———. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100: 763–801.

Rudebusch, Glenn D. 1998. "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review* 39: 907–31.

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50, no. 1: 3–50.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2011. "Measuring Mismatch in the U.S. Labor Market." Federal Reserve Bank of New York.

Sargent, Thomas J. 1989. "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy* 97: 251–87.

Sargent, Thomas J., and Christopher A. Sims. 1977. "Business Cycle Modeling without Pretending to Have Too Much A-Priori Economic Theory." In *New Methods in Business Cycle Research,* edited by C. Sims and others. Minneapolis, Minn.: Federal Reserve Bank of Minneapolis.

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96, no. 1: 54–81.

Smets, Frank, and Raph Wouters. 2007. "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97, no. 3: 586–606.

Stock, James H., and Mark W. Watson. 2008. "NBER Summer Institute Minicourse 2008: What's New in Econometrics—Time Series, Lecture 7: Structural VARs." Cambridge, Mass.: National Institute for Economic Research. www.nber.org/minicourse_2008.html.

———. 2009. "Forecasting in Dynamic Factor Models Subject to Structural Instability." In *The Methodology and Practice of Econometrics: Festschrift in Honor of D. F. Hendry,* edited by Neil Shephard and Jennifer Castle. Oxford University Press.

———. 2011. "Dynamic Factor Models." In *Oxford Handbook on Economic Forecasting,* edited by M. J. Clements and D. F. Hendry. Oxford University Press.

Comments and Discussion

COMMENT BY

ALAN S. BLINDER   James Stock and Mark Watson, as skilled a pair of time-series econometricians as the profession boasts, ask in this paper how the 2007–09 recession differed from other U.S. postwar recessions. Their answer is provocative: not much, actually, it was just *bigger*. They conclude that "the 2007–09 recession was the result of one or more large shocks, that these shocks were simply larger versions of ones that had been seen before, and that the response of macroeconomic variables to these shocks was almost entirely in line with historical experience."

Provocative? Yes. I must admit to being provoked. For starters, virtually every previous postwar recession was caused by either tight monetary policy, an oil shock, or both.[1] This one was not. Instead, it appears to have been caused by the bursting of a gigantic home-price bubble and an even more gigantic fixed-income bubble, which together led to the near-collapse of a jerry-built financial system, causing massive wealth destruction and severe impairment of the economy's credit-granting mechanisms. It looks more like a Reinhart-Rogoff recession, the first in U.S. postwar history, or better yet, a Minsky recession.[2] Furthermore, the Federal Reserve quickly (in December 2008) encountered the zero lower bound (ZLB) on nominal interest rates, forcing it to rely on a wide variety of "unconventional" monetary policies. Finally, long-term unemployment soared to levels not seen since the 1930s. In short, the years following 2007Q4 did not look

---

1. Among the many papers that could be cited, see Bernanke, Gertler, and Watson (1997).

2. By a Reinhart-Rogoff recession I mean a recession caused or prolonged by the harmful balance-sheet effects of a financial crisis (Reinhart and Rogoff 2009); by a Minsky recession I mean a recession caused by the (inevitable) bursting of an asset bubble after excessive speculation (see, for example, Minsky 1992).

like your father's recession—more like your grandfather's (or your great-grandfather's, if you are very young).

Let me start by reviewing why the conventional wisdom sees the 2007–09 recession and its aftermath as something different. Then I will take up Stock and Watson's challenge to this wisdom.

THE CONVENTIONAL WISDOM The U.S. economy entered 2007 with incredibly high leverage, virtually everywhere. Leverage is not something new. But as recently as the last mega-recession, in 1982–83, total debt in the United States was only about 150 percent of GDP. By the end of 2007, it was about 350 percent of GDP. High leverage, of course, spells high vulnerability to shocks. That vulnerability, in turn, was exacerbated by a variety of complicated (and mostly unregulated) financial linkages—through complex derivatives, a variegated shadow banking system, and more—that created a house of cards in the years leading up to the crisis. These novel financial developments sound like "slope things" to me—aspects of the economy that would likely change impulse response functions—as opposed to "intercept things" that only change levels of variables. This financial house of cards, by the way, was poorly understood, which makes the standard assumption of rational expectations something between dubious and ludicrous.

Yet more "slope things" relate to monetary policy. When the Federal Reserve hit the ZLB in December 2008, that presumably reduced the power of monetary policy. Certainly it ended the ability to use the federal funds rate as a policy instrument, even though linear equations in standard models call for the funds rate to go negative. Stuck at the ZLB, the Federal Reserve turned to new, untested weapons like emergency lending facilities, large-scale asset purchases, and forward guidance—policies whose "multipliers" are mostly unknown.

Finally, as mentioned, the share of long-term unemployment (spells over 26 weeks) in total unemployment rose to over 45 percent, having never before topped 26 percent in postwar history. Might all those long-term unemployed people affect the responses of measured employment and unemployment to stimulative policies?

For all these reasons and more, there is a strong *a priori* case that empirical models based on historical data might be expected to perform poorly after 2007. Before we jump to that conclusion, however, let us note several other factors pointing in the opposite direction, which tend to support Stock and Watson's claim that this recession wasn't different, just bigger.

Bubbles have burst before, although this one was the whopper. We have also regularly witnessed deterioration in underwriting standards and other

signs of irrational exuberance in lending during past booms. The financial zaniness of 2004–07 was not the first "Minsky moment" in postwar U.S. history—though it looked more like a Minsky quinquennium. Nor was what followed the Lehman Brothers bankruptcy the first banking crisis. The sheer size of the 2007–09 recession was unprecedented: real GDP declined by nearly 5 percent; even nominal GDP fell. The decline in employment (and the rise in unemployment) also appeared to be unusually large, even given the miserable GDP performance, and the bounceback of employment after the trough seems unusually slow. But maybe these are all just exaggerated versions of what we have experienced in past recessions: normal reactions to shocks that, although abnormally large, are not qualitatively different.

This last thought leads to a philosophical question that is, in some sense, at the heart of Stock and Watson's analysis: When does a *quantitative* change become so large that it becomes a *qualitative* change? For example, the earth's tectonic plates are always moving, but somehow, earthquakes are different. Years ago, Thomas Sargent (1982) called our attention to the unusual behavior around "The Ends of Four Big Inflations" in Central Europe in the 1920s. There is certainly no apparent Phillips curve trade-off, and no "stickiness," when the inflation rate drops by hundreds or thousands of percentage points within a few months, as it did then. And Robert Shiller (2008) has pointed out that the U.S. housing boom and bust of the 2000s was unlike anything seen in the nation's history dating back to 1890. So was the recent bubble *sui generis,* or just "normal but bigger?" Stock and Watson implicitly argue for the latter. I wonder.

And while we are thinking this way, is the bursting of a bubble a "shock," as we conventionally use the term? After all, we all knew with near certainty that the housing bubble was going to burst; the only question was when. If so, did the ε in the home-price equation still have a mean of zero after, say, 2005? Indeed, were Christopher Sims, who does not allow such words, not the other discussant of this paper, I might be tempted to ask whether bubbles are *exogenous* or *endogenous* variables.

THE STOCK-WATSON VIEW The essence of Stock and Watson's economet-ric methodology is as follows. Back in ancient times, economists used to estimate giant "structural" macroeconomic models that could be used to derive reduced forms like

(1) $$Y = X\beta + \varepsilon,$$

where Y is a vector of "endogenous" variables explained by the model, X is a vector of "exogenous" variables not explained by the model, and ε is a

vector of error terms, or "shocks." One problem was that Y and X might be very long vectors, making β a truly gigantic matrix, with more parameters than one can reliably estimate. The factor analysis approach is an attempt to economize on parameters, starting from the observation that one can always find a much smaller set of variables, Z, such that

$$(2) \qquad X\beta = Z\Gamma + u,$$

where u is another error vector. How well equation 2 fits the data is an empirical question whose answer will depend, among other things, on the number of Z variables—the "factors." Using equation 2, one can rewrite equation 1 as

$$(3) \qquad Y = Z\Gamma + \varepsilon^*.$$

Whereas X (and thus β) may be huge, Z (and thus Γ) will be of manageable size. In Stock and Watson's particular application, Z is only six-dimensional. The operational question is how much information is lost in going from equation 1 to equation 3; that is, how good an approximation equation 2 is.

This formalism is mathematically valid, of course, but let me point to three weaknesses. The first is what I call the *reification fallacy*: One tends to treat the εs in either equation 1 or equation 3 as actual *things*—"shocks"—when they are really *deviations from conditional expectations* ("error terms"). To be sure, there are such things as genuine unexpected *shocks,* like surprise movements in oil prices or in monetary policy. But what economists typically call a "consumption shock," for example, is just the error term in the consumption equation; it expresses, among other things, our inadequate understanding of consumer spending. This distinction is highly relevant to the "was it just bigger?" question. For example, when several 5- or 6-standard-deviation *shocks* are observed, are those just unusually large error terms, or do they signify that something virtually unprecedented happened?[3] I lean toward the latter.

Second, although moving from equation 1 to equation 3 is legitimate, both algebraically and statistically, the Zs and Γs cry out for interpretation. Often that interpretation is hard to give, or simply not given. In the old days, the Xs in equation 1 had clearly understandable names like "exports" or "government purchases." So, for example, $\partial Y_1/\partial X_1$ might be interpreted as the multiplier effect of higher exports on real GDP. In describing their

3. Stock and Watson's table 5 contains six shocks larger than 5 standard deviations; three of them are larger than 8 standard deviations.

factor analysis methodology, by contrast, Stock and Watson write, "arbitrary normalization means that the individual factors do not have a direct economic interpretation." I am, no doubt, an old fuddy-duddy, but this strikes me as a drawback. If, in the old methodology, our computations found that $\partial GDP/\partial exports$ was 10 or -1, we would know immediately that something was wrong. We have no such intuition about $\partial GDP/\partial F_1$.

The third problem is obvious, but it is also important in the context of deciding whether some phenomenon is "new" or just "bigger." If something has not been experienced before, it obviously will not be in a statistical model based on real data. For example, oil shocks were nowhere to be found in pre-1973 macroeconomic models, although James Hamilton (1983) later taught us that they should have been there all along. How, then, can a purely statistical model—as opposed to observation, common sense, or an *economic* model—tell us whether some unusual development is "new" or just an unusually large deviation from historic norms?

With all that said, Stock and Watson's factor analysis model performs surprisingly well. For most of the major macroeconomic variables, such as GDP, employment, and their main components, their factor analysis model estimated on *pre*-2007Q4 data captures the *post*-2007Q4 data amazingly well. (See many of the panels in their figure 2.) I was impressed.[4]

But who ever expressed the view that, say, consumption behaved abnormally relative to its major determinants (disposable income, wealth, and so forth) during this period? The Stock-Watson factor analysis model misses badly more or less where events lead you to think any model would. Home prices, bank lending, the federal funds rate, the monetary base, and long-term unemployment are some examples (again from their figure 2). I do not say this to criticize the authors; those variables are awfully hard to "get right" during the recession and its aftermath. But it does lead me back to the thought that the 2007–09 recession and subsequent recovery really were different, not just bigger.

Perhaps the gentlemen protest too much.

REFERENCES FOR THE BLINDER COMMENT

Bernanke, Ben S., Mark Gertler, and Mark Watson. 1997. "Systematic Monetary Policy and the Effects of Oil Price Shocks." *BPEA,* no. 1: 91–142.

Hamilton, James. 1983. "Oil and the Macroeconomy since World War II." *Journal of Political Economy* (April): 228–48.

---

4. Maybe more than I should have been. Some Panel members suggested during the general discussion of the paper that one of the six factors must have closely resembled real GDP.

Minsky, Hyman P. 1992. "The Financial Instability Hypothesis." Working Paper no. 74. Jerome Levy Economics Institute of Bard College.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Sargent, Thomas J. 1982. "The Ends of Four Big Inflations." In *Inflation: Causes and Effects,* edited by Robert E. Hall. University of Chicago Press for the National Bureau of Economic Research.

Shiller, Robert J. 2008. *The Subprime Solution.* Princeton University Press.

COMMENT BY

CHRISTOPHER A. SIMS   This paper by James Stock and Mark Watson is an exercise in descriptive statistics, attempting to provide insight into what was surprising about the Great Recession compared with historical statistical patterns. Economists are attempting to revise their theories and empirical models to account for what happened in 2007–11, so this paper's systematic examination of what happened, using a wide range of economic time series, is valuable.

The paper has two themes: some things about the Great Recession were very different from historical patterns, and some were not. The story it tells, with some qualifications, is that the Great Recession was characterized by big "shocks" (forecast errors in a linear model), but that those shocks fed through the dynamics of a linear model much as one would have expected from historical patterns. The paper identifies the recent phenomenon of "jobless" recoveries as differing from historical patterns, but in a smoothly trending way that was visible well before the Great Recession.

That there were big shocks is clear, and it is interesting to see when they occurred and in which variables. The downward bend in trend employment growth that the authors find looks statistically convincing, and the paper informally suggests plausible reasons for it. The case for the linear dynamics having been stable is less convincing to me, however, and there are aspects of the data's behavior that have been unusual and that are not captured in pre-2008 economic models or in the models that this paper fits to the data.

The biggest statistical surprise in the recession was the size of the forecast errors in the linear model, both during the 2008Q4 crash and to some extent preceding it. The paper does not actually display any such forecast errors, but table 5 shows estimated surprise components of 11 variables, constructed from the authors' dynamic factor model. It would be easier to interpret the results if these were actual forecast errors in the variables

listed at the top of the table, but as the paper's figure 2 shows, the variables are fairly close to the same thing as their "common components," so the table comes fairly close to showing us forecast errors in these variables. The errors are scaled by the standard deviation of forecast errors over 1960–2007Q3, so those larger than 4 in absolute value—in productivity, housing starts, oil prices, the VIX, and the TED spread—are extremely unusual by historical standards. If the disturbances were normally distributed, even those larger than 3 would be unlikely to have occurred in the historical sample, but the data are not normal. Three-sigma errors occurred fairly often in the sample, but four-sigma and larger errors did not. In fact, all the series with over-four-sigma errors had no disturbances of that magnitude in the sample period. (These conclusions are all based on data related to table 5, extending it to earlier periods, that the authors made available to me.)

That there were large unforeseen changes in economic time series in this period is, of course, not news. It is good to have the size of the surprise quantified, however, and to observe where the surprises concentrate: Except for the 2008Q4 productivity shock (about which more below), each of these extremely large surprises was in a variable related to housing or financial markets. That these variables showed large unexpected changes is again not news, but their size relative to historical norms is worth noting.

Perhaps more unexpected is that for GDP, consumption, and investment there were large disturbances, but they were not larger than had been seen during the 1960s and 1970s. This raises an interesting question that the paper does not explore: did the large shocks in financial variables feed through into large predicted changes in the nonfinancial variables, and did those effects lead to more accurate forecasts?

My curiosity aroused, I used monthly data from the authors' database on the 3-month Treasury bill rate, the 3-month London Eurodollar deposit rate, and logs of industrial production, employment, oil prices, and the personal consumption expenditures deflator to estimate a vector autoregression (VAR) on data from 1971 through September 2007. (I used 13 lags and an improper prior shrinking toward persistent behavior of the data.) If, as seems likely to me, the paper's six factors can be well approximated by linear combinations of current and lagged values of these six variables, one can use the VAR to gain insight into how the paper's model works.

The VAR confirms that the Eurodollar rate (which, combined with the Treasury bill rate, can reproduce a version of the TED spread) has substantial power in explaining industrial production over medium to long horizons. When the Eurodollar rate is excluded from the VAR, forecasts

conditioned on data through September 2008 (although still with coefficients estimated through September 2007) are considerably worse. The Eurodollar rate variable allows the VAR to underestimate the decline in output and employment through mid-2009 by substantially less.

As can be seen from my figures 1 and 2, however, the VAR residuals contrast with the findings reported in the paper's table 5 in some respects. One is that the very large forecast errors are spread evenly across variables in the VAR, not concentrated as in the authors' model in its one financial-stress variable LIBOR. This could happen because the VAR does not include as rich a set of financial variables, but in my view it is more likely a reflection of the difference in time units. What happened to output, employment, and the PCE deflator in September 2008 was extremely large and sudden by historical standards. The suddenness is partly smoothed away by the use of quarterly rather than monthly data. The large positive shock to productivity shown in the paper's table 5 probably reflects the complicated and rapid changes in output and labor markets during this period. No positive shock to productivity shows up in the VAR residuals, and the authors in a personal communication to me showed that the positive shock occurs only in the projection of productivity on the factors, not in the productivity series itself. Most of the surprises in table 5 correspond well with sharp movements in the series labeling the columns, but not this one.

Many of the monthly series show VAR residuals that are both large and oscillating in sign during the final 3 months of 2008 and the first half of 2009. During this period the large shocks were not simply feeding through the usual dynamics. The usual dynamics were producing large error after large error. After this period, however, the residuals return to a size commensurate with historical norms. My conclusion is that the notion that the crisis consisted merely of large shocks feeding through the usual dynamics is somewhat misleading. The usual dynamics did not explain what was going on for several months around the peak of the crisis. During that period existing linear models may have been unhelpful. But after that period, starting from a new, depressed economic state but with the financial markets stabilized, models fitted to history began tracking reasonably well again.

My figure 1 brings out another point that is implicit in the paper's methods but could be further emphasized. The historical data before the end of 2007 already showed clear evidence that large outliers and sustained periods of unusually high or low volatility are recurring phenomena. All the statistical tests that the paper applies carefully avoid assuming normality of residuals or even constant variances, sacrificing power because the stronger assumptions would be clearly counterfactual. If one is looking for what the

Figure 1. Standardized Monthly Residuals for 1970–2011 from a Vector Autoregression Fitted to Data through September 2007

3-month Treasury bill yields

Standard deviations

Employment[a]

Standard deviations

3-month London Eurodollar deposit rate

Standard deviations

Oil price[a]

Standard deviations

Industrial production[a]

Standard deviations

Personal consumption expenditures deflator[a]

Standard deviations
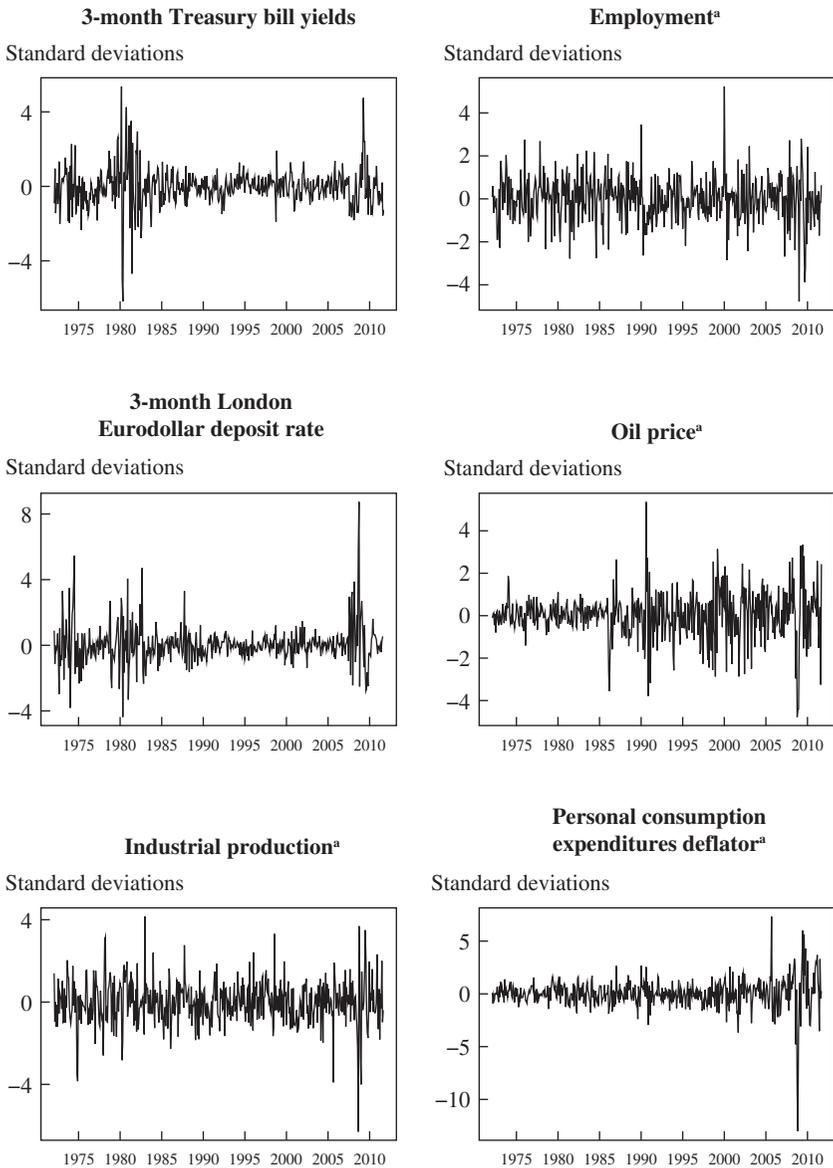
Source: Author's calculations. See the text for details.
a. Variable is in logarithms.

Figure 2. Standardized Monthly Residuals for 2007–11 from a Vector Autoregression Fitted to Data through September 2007
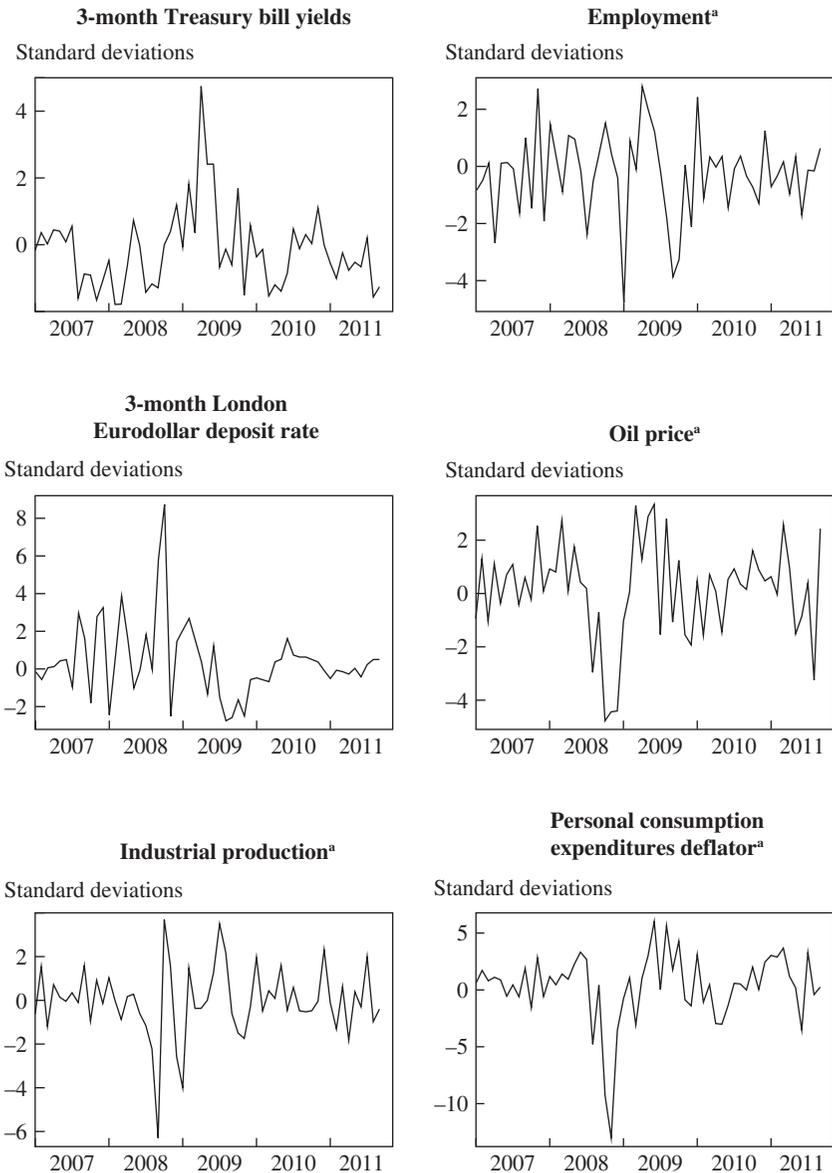
Source: Author's calculations. See the text for details.
a. Variable is in logarithms.

standard macroeconomic models have been missing, this is a good place to start. We should be modeling the evolution of volatility and its potential interaction with mean dynamics, not treating these phenomena as "nuisance parameters" to be worked around. Our statistical descriptive models should go beyond linear models, in other words, and our theoretical models should be required to provide interpretations for the results. The interaction of financial market malfunction with the macroeconomy may be a good part of what generates the outliers and time-varying volatility that we observe.

The authors were among the first to point out the value of interest rate spread variables in forecasting. In their seminal 2002 paper they included a number of spread variables, but not the TED spread or a Eurodollar rate. In this recent crisis the Eurodollar rate has done particularly well in after-the-fact forecasting models, as it measures distress in interbank lending, which was central in this crisis. It was useful before the crisis as well, as otherwise models estimated through late 2007 would not have detected its effects. But the fact that it plays a prominent role in the paper's model does reflect what economists have already learned about the deficiencies in precrisis models. Measures of financial distress are important, and we have been sifting through candidates for measuring them better. Financial institutions and regulations are continually changing, however, so that the markets where we need to look for measures of stress will probably continue to shift. This suggests a reason for caution in interpreting this paper's results as showing big shocks feeding through stable dynamics.

My conclusions about the substance of the paper's results are that it is correct to point to the continuing usefulness of models fitted to historical data in the wake of the crisis, but that to suggest that this means the crisis was "just large shocks" is misleading. Linear model dynamics did not work well during the peak of the crisis, and linear models turn a blind eye to what should be one of our most important scientific tasks, that of understanding recurrent instances of high volatility in macroeconomic and financial time series.

I turn now to some comments on the paper's methods. The paper uses the principal-component dynamic factor model setup that the authors developed and have used before. The approach requires stationary data, and economic data are not stationary. Accordingly, the authors prefilter all series to make each one plausibly stationary, first-differencing real variables and second-differencing prices. My view of the econometric literature on cointegration is that its complicated prescriptions for frequentist inference—which would be totally impractical for the model of this

paper—are misguided. But the literature begins from an important observation: when cointegration may be present, simply getting rid of nonstationarity by differencing individual series so that they are all stationary throws away vast amounts of information and may distort inference. That is why the VAR that I used above to interpret the paper's results was estimated in levels, allowing the model dynamics to account for any nonstationarity.

In addition to this differencing, the paper removes "trend" through another layer of filtering, then applies factor analysis to the residuals from these two layers of preprocessing. Any inference that is carried out ignores uncertainty about the prefiltering and about the factors themselves. As mentioned above, all the statistical tests are in a form that, in a large enough sample, would be robust to nonnormality. Because of these layers of data processing and of the authors' reluctance to present an explicit model of nonnormality and time-varying volatility, it would not be reasonable to ask them for error bands around the statistics they develop. Perhaps most important, their plots of out-of-sample forecast and actual values of employment growth (figure 4 in the paper) are presented only as point forecasts. One can see that the model, in most but not all recessions, gets the overall growth rate of employment about right, but there are errors, and no standard exists by which to judge whether these errors are big or small. This is one main reason that I prefer to work with an explicit probability model of the data, which becomes testable because it attaches uncertainty measures to its forecasts.

The preprocessing also raises some doubts about the result that jobless recoveries are explained by a downward bend in the trend rate of growth of employment. Because the trend is estimated with a two-sided filter that uses data for several years ahead of the current date, the precrisis trend could actually be affected by crisis-period data.

The complexity of the dynamic factor model structure could in principle be providing better forecasts and more accurate modeling of the macroeconomy. This paper does not provide any direct evidence on whether this approach is an advancement over one that would directly model the time-series behavior of, say, a dozen major standard aggregate series, explaining the remaining 150 or so as functions of these central aggregates. My reading of the previous literature on these methods is that it does not provide evidence on this point either. This matters, because the task of explicit probability modeling is much harder if it requires extraction of unobservable factors from hundreds of series.

I have not commented on the paper's attempts to give behavioral interpretations to its factors by correlating them with external variables. I find

this exercise completely unconvincing. The external variables have themselves in most instances been constructed by modeling macroeconomic aggregates, many of which are included in the factor model or have close analogues in it. The intuitive interpretation of this procedure is similar to that underlying the projections of individual variables on factors that generated the results in table 5. Some variable or group of variables to which one can give a name is regressed on the factors, and the resulting linear combination of factors is given a name. That may help in interpreting the six nameless factors, but it is in no sense "structural" estimation. The authors do not even provide a structural model of the data within which these estimates would be valid estimates of something structural.

The paper represents hard work and has produced much food for thought. Like most good empirical work, it required making decisions about how to analyze and present data (that is, "assumptions") that provide plenty of targets for critics. But despite my having taken aim at some of these targets, I hope the paper will stimulate more work along this line.

REFERENCE FOR THE SIMS COMMENT

Stock, James H., and Mark W. Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20, no. 2: 147–62.

GENERAL DISCUSSION   William Brainard noted that the authors' conclusions depended heavily on how they distinguished between changes in structure and shocks. That distinction, in turn, depended on how they detrend and how they weight the roughly 200 time-series variables. Their estimate of a variable's trend was approximately the same as its average rate of change over a centered window of plus or minus 30 quarters. If a shock is large and persistent, as arguably some were in the 2007–09 recession, this procedure may improperly allocate a significant fraction of a shock to the trend. Brainard cited as an example the case of housing starts, which were relatively steady at a high level until the bubble burst, declined rapidly starting in 2005, and have remained low ever since. When the authors' detrending procedure is used, much of the large drop in housing starts is likely to appear as a decrease in trend. The dramatic drop in starts does not average out; the shocks in the window covering 2007–09 go in one direction. The same danger of confounding trend and cycle, Brainard continued, arises in the case of employment growth, where the authors attribute most of the recent slow recovery to a slowdown in trend employment growth.

Brainard also reasoned that the results could be sensitive to the authors' use of unweighted percentage changes for the real variables. As a consequence, their results could reflect large percentage changes in variables that are relatively unimportant to aggregate output, while underweighting variables that are more important but have smaller percentage changes. Principal component analysis, which the authors use to compute the factors, is not scale independent, Brainard noted. He wondered what the results would have been if the authors had weighted changes by their importance to the overall economy or had conducted their analysis in levels.

Finally, the authors' finding that the structural dynamics of the pre-2007 recessions matched those of the most recent one puzzled Brainard, since the most recent recession was so deep and the current levels of variables like housing starts and unemployment are so different from what was observed during previous recoveries. In the case of housing starts, the authors' scatterplot makes clear that the errors using the dynamic factor model are autocorrelated, which is inconsistent with the model's assumptions. Brainard suggested that such autocorrelation of errors might explain similar discrepancies between the observed levels of other variables and the predicted levels implied by the authors' model.

Refet Gürkaynak observed that after every big macroeconomic event, economists feel pressure from the public and policymakers to look at that event as something entirely new and unique and to forget all lessons from previous experience and research. He saw this paper as an important counter to that kind of thinking, because it showed that economists have accumulated some wisdom over the years that remains relevant from event to event and could be used to give policy advice even in the face of new challenges.

Gürkaynak was troubled, however, by the fact that a large financial shock was driving the results of the authors' model for the recent recession. He interpreted financial asset prices as forward-looking information aggregators, which measure whatever is missing in the rest of the economic model. The fact that the model required a large financial shock to explain the Great Recession data meant that there are probably important variables that were observable over the period leading up to and during the recession but are omitted from the model and instead show up indirectly as the large financial shock.

Olivier Blanchard argued that it is in the nature of dynamic factor models to explain aggregate variables fairly well, because many of the series are closely related to each other and therefore highly correlated. For example, some of the authors' series might be different components of industrial production, which the industrial production index, an aggregate of these, will explain quite well.

Blanchard did not take the fit of the linear model as evidence that structural relationships in the economy were truly linear. On the contrary, he thought Stock and Watson did not provide a way to test empirically for nonlinear relationships, but he felt confident that such relationships exist. The effect of the interest rate on debt dynamics, for example, depends on whether the national debt is 50 percent of GDP or 100 percent, as the recent European experience shows. The effect of a 1-percentage-point drop in the rate of GDP growth on loan performance, as another example, depends on whether the GDP growth rate starts out at 4 percent or at 1 percent. Blanchard suspected that coming up with appropriate ways to measure these relationships would be hard but could be done.

Martin Baily said that given the unusual nature and size of the shocks that had provoked the Great Recession, he was surprised by the extent to which Stock and Watson found the shape of the recession and subsequent recovery to match that of previous recessions and recoveries. The recent recession had featured a strong inventory cycle, as had previous ones, for example. And a financial accelerator effect appeared to have depressed investment, again as in previous recessions.

What Baily found most puzzling was the performance of the labor market, which he thought had suffered more initially and recovered more slowly than in previous recessions even after accounting for demographic changes. He was therefore surprised that the dynamic factor model was able to match employment data in the recession and recovery as well as it did. Baily thought that cross-country data might shed some light on the poor U.S. labor market performance. He noted that patterns in GDP growth and employment growth varied widely across countries during and after the Great Recession. Germany and the United Kingdom both suffered larger drops in GDP but experienced smaller declines in employment than the United States. Spain, on the other hand, experienced a similar drop in GDP as the United States but an even larger employment decline. The other aspect of the Spanish economy that paralleled the U.S. experience was its large housing boom and subsequent collapse of residential construction. That parallel led Baily to wonder whether the exceptional labor market decline in both countries could be traced to the collapse in housing construction. The theory might explain why Stock and Watson's model could explain the collapse in employment, since their analysis includes housing market data.

Frederic Mishkin, like Blanchard, thought that modeling nonlinear features of the business cycle was very hard but worth further study. The literature has long noted the existence of financial disruptions and financial

accelerators as important factors in recessions, he said, but these phenomena had not been incorporated in the dynamic stochastic general equilibrium models used for policy analysis. He wished especially that models could better incorporate the extreme nonlinearities that are constantly present in the financial sector and worsen during recessions. Some of these nonlinearities could even independently help trigger recessions, as he thought was true of Enron's crash in late 2001 and the recession that soon followed. For the most recent recession, he saw the crash of Lehman Brothers as an important source of nonlinearity.

Robert Hall described a related exercise he had recently carried out to examine the causes of the recession. Using a dynamic, nonlinear macroeconomic model, he computed the paths that two driving forces—financial friction and deleveraging—must have taken to explain the pattern of unemployment and investment observed during the downturn. He found that financial friction was a large and persistent factor in explaining unemployment and investment, whereas deleveraging was an important factor early on but diminished in importance quickly. The same exercise would have been easy to carry out for earlier business cycles using the same model, the only difference being that the movements of the variables would have been smaller. That the same model and driving forces could explain patterns of unemployment and investment across different recessions was unsurprising, since they did so by design.

Because Hall saw Stock and Watson's exercise as similar to his own, he found their results mostly unsurprising, with the possible exception that they could explain a large number of time series well using a much smaller number of factors. However, he understood it to be a well-known feature of dynamic factor models that only a few factors could explain the movement of many variables.

Linda Tesar wondered how well the authors' model could explain imports and exports, since both dropped significantly during the recession even though the exchange rate changed little. She also suggested that another test of the model might be to test how well it fits the data in other countries.

Justin Wolfers thought the authors had neglected to highlight one of the most striking findings of the paper, namely, that numerous variables that had previously been identified by macroeconomists as exogenous instruments were in fact highly correlated with each other. In one of their tables, for example, the fiscal shock identified by Christina Romer and David Romer in 2010 exhibited a −0.8 correlation with the monetary policy shock identified by the same authors in 2004. Wolfers noted that if labor economists somehow discovered that the typical instruments for educational

attainment—quarter of birth, Vietnam draft lottery number, and distance to college, for example—were in fact highly correlated, they would regard the finding as enormously destructive to most of what modern labor microeconomics had achieved. He wondered, then, what this correlation of supposedly instrumental variables meant for macroeconomics. Wolfers also found it puzzling that the authors' factors seemed to explain GDP and some other macro time series too well given the margin of error with which those series are measured.

John Driscoll offered a possible insight into why one of the series, commercial and industrial loan volume, fit less well than some others. Early in a recession, business loans often rise before eventually falling, because firms are using previously unused parts of credit lines that had been approved before the recession started. The recent recession featured a particularly large run-up and a subsequent crash in commercial and industrial loans. One could observe this effect more directly by examining measures of *unused* loan commitments from banks' call reports, which fell at the start of the recession as businesses used up their available credit.

Valerie Ramey suggested a caveat to the authors' finding that an oil shock was an important contributor to the start of the recession. Previous research had found that the effects of oil price shocks on the U.S. economy had declined since the mid-1980s. However, in a 2010 paper, she and Daniel Vine had found that the appearance of a structural shift in the effect of oil shocks could be traced to the removal of price controls on oil after the 1970s, which had caused severe misallocation. With price controls absent, an increase in the price of oil in 2008 ought to have had a smaller effect on the economy than a comparable shock in the 1970s.

Ramey also commented on the difficulty of separating shocks from long-run trends, noting that the latter may actually help drive what we think of as the business cycle. In unpublished work, using a standard real business cycle model, she had modeled the effects of an anticipated change in the growth rate of the labor force and found that decreases in that rate in the 1930s had led to decreased productivity and investment. Similarly, John Maynard Keynes and Alvin Hansen had suggested that the cutoff of immigration in the 1920s, which slowed labor force growth, was a factor in the Great Depression. Ramey had also found that, in the 1940s, an increase in the labor force growth rate led to increased productivity and investment even without a World War II shock in the model.

Christopher Carroll thought the paper was helpful for focusing economists' attention on the shocks that had led to the recession. However, he thought the most credible method of calculating the recession-causing

shocks might be to use a model that was published just before the recession. David Wilcox reported that several of his colleagues, Hess Chung, Jean-Phillipe Laforte, David Reifschneider, and John Williams, had completed an exercise similar to the one Carroll had suggested; their paper was forthcoming in the *Journal of Money, Credit, and Banking.* Before the Great Recession, using the 2007 version of the Federal Reserve's main macroeconometric model, the FRB/US model, Reifschneider and Williams had found that the likelihood of a multiyear zero-lower-bound event was so low as to be practically a statistical impossibility. The forthcoming paper uses a range of other models and considers a wider spectrum of sources of uncertainty. With those adjustments, recent events are seen as less remotely improbable, but still relatively unlikely. On this basis, Wilcox was surprised that Stock and Watson's paper seemed to suggest that their model could account for basic features of the recession.

To Ricardo Reis, the fit of the model to the post-2007Q4 data was useful for demonstrating that over the course of the recent recession and recovery, the six-factor model was not missing a major seventh factor. Reis highlighted a footnote in the paper in which the authors specify the algebraic manipulations they used to estimate the post-2007Q4 factors and to compute the common component of the macroeconomic time series in the post-2007Q4 period. There they explain that they use principal component analysis to find the six linear combinations of the series that best account for the data over the 1959–2007Q3 period. These linear combinations are described by the factor loadings matrix, $\hat{\Lambda}^{59-07}$. The six factors are then estimated for the post-2007Q4 period by applying the transpose of the factor loadings matrix to post-2007Q4 data for the macroeconomic series. Finally, the factor loadings matrix is applied to the six estimated factors to predict the same series in the post-2007Q4 period. The key point for Reis was that, with this methodology, the post-2007Q4 factors are "old" in the sense that they are based on the pre-2007Q4 linear combinations of the series. As the footnote explained, if there were a new factor, the space spanned by the factors would change, so the new factor would not be spanned by the post-2007Q4 estimated factors.

Reis likened this process to a much simpler exercise: Imagine being given data on durable goods output and nondurable goods output for a number of years. These two series sum to total goods output. Then suppose one is given total goods output for the next few years and asked to predict durable goods output and nondurable goods output during that period. The strength of the prediction will depend on how well one is able to capture the relationship among durable goods, nondurable goods, and total output

over the period for which data on all three series are available, and whether this relationship continues to hold in the period over which the prediction is made.

Reis drew from the authors' analysis a much weaker claim than some other panelists had suggested. The exercise did not, he argued, imply that the world did not change in 2007. The six factors may have undergone larger innovations than in past decades, or become more stochastically volatile, but such changes would still be consistent with the authors' finding, so long as no new type of shock had emerged.

Reis saw the authors' "no new shock" finding as consistent with theorizing by other economists trying to explain the sources of the financial crisis. That work often focuses on just one or two types of structural shocks—such as preference shocks, technology shocks, or monetary policy shocks—and abstracts away other aspects of the economy. Thanks to Stock and Watson's work, Reis thought, these researchers could rest assured that they are not ignoring some unidentified "animal spirit" or other type of shock that could undermine their models.

Finally, following on Wolfers' point, Reis thought that the correlations between variables that had previously been identified as instrumental show that economists have limited understanding of what structural shocks are driving the factor innovations in Stock and Watson's model.

Wendy Edelberg noted that, at the trough of the recession, if policymakers had been able to predict the path of GDP growth over the next few years, they presumably would have enacted different policies. She wondered, then, whether the authors' analysis shed any light on what policymakers should have done during the recession.

Christopher Sims highlighted the fact that Stock and Watson's results were driven, in part, by large deviations from historical norms in the interest rate and in monetary aggregates. He thought it was misleading to acknowledge these large historical deviations but ignore the fact that, in their absence, other macroeconomic variables would have behaved very differently. He saw Stock and Watson's inability to predict what would have happened in the absence of significant policy intervention as a limitation of their analysis.

Sims analogized the issue to tracking the temperature in one's kitchen from day to day. On a typical day, the temperature rises while dinner is cooking and then falls. Suppose one day a fire started while dinner was being prepared and a fire extinguisher was used to put the fire out. The time path of temperature in the kitchen would look relatively normal, but it would be incorrect to say nothing unusual had happened, because had the

fire extinguisher not been used, the temperature in the kitchen would have developed very differently that evening.

Michael Kiley characterized the authors' work as a tracking exercise rather than a prediction exercise. He did not find it surprising that their factors were able to track macroeconomic aggregates well. He saw their principal component analysis as identifying an "economic activity" factor, which, much like the Federal Reserve Bank of Chicago's National Activity Index, could track GDP well using a linear combination of other macroeconomic time series. Applying Okun's Law, one could also use the same factor to track the unemployment rate fairly well.

Kiley thought the authors were correct to identify the financial shock as an important driver of the recession, and correct to say that, conditional on the large shock to financial prices, the macrodynamics that followed were not so surprising. However, their model did not incorporate a rich enough picture of the financial system to determine whether the forces that created or amplified the large financial shock were surprising or unusual or nonlinear. The panic related to mortgage-backed securities and the collapse of three investment funds held by the French bank BNP Paribas, for example, were unusual events that seemed to have induced large movements in financial markets but were not captured by Stock and Watson's model.

Responding to the discussion, James Stock agreed with Brainard that his and Watson's detrending method may not always distinguish well between shocks and trends, especially near the end of certain data series. He and Watson had run sensitivity checks that gave them confidence in their results, but they thought that more work needed to be done in identifying trends in macroeconomic series and that researchers would do well to study trend identification more closely.

Replying to Wolfers, Stock felt that macroeconomists' work on finding exogenous instruments was very constructive and well worth continuing. He did not view the correlations between instruments as a setback but rather as a source of new questions to pursue.

Stock also thought that Reis's interpretation of his and Watson's findings was essentially correct. To further clarify, he explained that, in the post-2007Q4 period, there did not appear to be any new factor driving correlation across the idiosyncratic errors in their model. And in linear algebraic terms, the space spanned by the factors also spanned the innovations in time series over the course of the recession. This seemed reasonable to Stock since the shocks driving the crisis, such as financial and uncertainty shocks, had occurred before on smaller scales.

Replying to Sims, Stock conceded that the results reflected only the net effect of financial shocks and countervailing policy interventions, measured on a quarterly basis. He and Watson had hoped to be able to examine these shocks at a finer level, but data limitations had prevented this.

Finally, Stock acknowledged that linear models like theirs were subject to substantial limitations and that nonlinearities could well be driving some patterns in the data. However, not all nonlinearities present problems; some should and do translate into shocks in their model. The zero lower bound, for example, is a nonlinear constraint that prevents the interest rate from falling below zero. In their framework this constraint translates into a contractionary monetary policy shock during the latter part of the Great Recession. However, nonlinearities that cannot be captured by shocks are not handled well by their model.