Journal of Financial Economics 119 (2016) 457–471

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Systemic risk and the macroeconomy: An empirical evaluation[☆]

Stefano Giglio[a,1], Bryan Kelly[a,*], Seth Pruitt[b,2]

[a] University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, United States
[b] Arizona State University W.P. Carey School of Business, P.O. Box 873906, Tempe, AZ 85287, United States

ARTICLE INFO

Article history:
Received 9 February 2015
Revised 7 April 2015
Accepted 21 April 2015
Available online 22 January 2016

JEL classification:
G10
E37
C32
C58

Keywords:
Systemic risk
Quantile regression
Dimension reduction
Macroeconomy

ABSTRACT

This article studies how systemic risk and financial market distress affect the distribution of shocks to real economic activity. We analyze how changes in 19 different measures of systemic risk skew the distribution of subsequent shocks to industrial production and other macroeconomic variables in the US and Europe over several decades. We also propose dimension reduction estimators for constructing systemic risk indexes from the cross section of measures and demonstrate their success in predicting future macroeconomic shocks out of sample.

1. Introduction

The financial crisis of 2007–2009 has made systemic risk a focal point of research and policy, and has established the financial sector as its center of analysis. The empirical side of the literature focuses on measuring distress in financial markets. This has produced a staggering variety of systemic risk proxies, many hoping to serve as an early warning signal of market dislocations like those observed during the crisis.

In this paper, we investigate how a buildup of systemic risk in the financial sector increases risks in the real economy. Specifically, we exploit the large set of existing measures of financial sector distress to quantify how fluctuations in systemic risk impact the probability of a macroeconomic downturn. We propose a new systemic risk index that efficiently aggregates recession-relevant information across the gamut of individual measures. We show that increases in the index are associated with a large widening

☆ We thank Lars Hansen for many helpful conversations about this project. We also thank our referee Laura Veldkamp, Tobias Adrian, Gadi Barlevy, John Cochrane, Doug Diamond, Rochelle Edge, Arvind Krishnamurthy, Nellie Liang, Sergio Rebelo, Amir Sufi, Amit Seru, Allan Timmermann, Jonathan Wright and seminar participants at the Federal Reserve Board, the Federal Reserve Bank of San Francisco, University of Chicago (Booth), Northwestern University (Kellogg), Yale University (SoM), Arizona State University (WPCarey), the Midwest Economic Association, the Western Finance Association, and the NBER 2013 Summer Institute EFWW group for helpful comments. We thank Xiao Qiao for excellent research assistance, Roger Koenker, Andrew Lo, Christian Brownlees, and Mark Flood for sharing Matlab code, and the Macro Financial Modeling (MFM) initiative for generous support. Data can downloaded from www.sethpruitt.net/GKPwebdata.zip.

* Corresponding author. Tel.: +1 773 702 8359.
E-mail addresses: stefano.giglio@chicagobooth.edu (S. Giglio), bryan.kelly@chicagobooth.edu (B. Kelly), seth.pruitt@asu.edu (S. Pruitt).
[1] Tel.: +1 773 834 1957.
[2] Tel.: +1 480 727 0762.

http://dx.doi.org/10.1016/j.jfineco.2016.01.010

of the left tail of economic activity. A one standard deviation increase in systemic risk shifts the 20th percentile of the industrial production (IP) growth shock distribution downward by more than 50%, from around −1.4% unconditionally, to −2.2% (annualized). In several occasions in US history (including during the recent financial crisis), the conditional 20th percentile was below −3%, twice as large as in normal times. In addition, we show that the systemic risk index also predicts reactions of policymakers: the 20th percentile of innovations in the Federal Funds rate drops by 60%, from −50 basis points (bps) to −80 bps.

Our analysis uses out-of-sample predictive quantile regression, which forecasts how specific features of the macroeconomic shock distribution respond to systemic risk. We argue that a quantile approach is appropriate for evaluating the potentially asymmetric and nonlinear association between systemic risk and the macroeconomy that has been emphasized in the theoretical literature.[3] These theories predict that distress in the financial system can amplify adverse fundamental shocks and result in severe downturns or crises, while the absence of stress does not necessarily trigger a macroeconomic boom. Quantile regression is a flexible tool for investigating the impact of systemic risk on the left tail of macroeconomic shocks, as opposed to focusing on their central tendency via least squares.

We examine 19 previously proposed measures of systemic risk in the US and ten measures for the UK and Europe. In building these measures, we use the longest possible data history, which in some cases allows us to use the entire postwar sample in the US. To the extent that systemically risky episodes are rarely observed phenomena, our long time series and international panel provide empirical insights over several business cycles, in contrast to much of the literature's emphasis on the 2007–2009 sample in the US.

We first investigate each systemic risk measure individually, asking whether or not it provides significant out-of-sample information about future macroeconomic shocks. Next, we ask whether it is possible to aggregate risk measures into a *systemic risk index* to enhance forecasting power. A naive way to do this is by including all the measures as separate right-hand-side variables in a multiple quantile regression. But we find that this approach has virtually no out-of-sample forecasting power. This is due to multiple quantile regression overfitting the sample data, analogous to well-understood problems of overfit in multiple least squares regression (e.g., Stock and Watson, 2006).

As an alternative, we propose dimension reduction techniques for a conditional quantile factor model. We show how these estimators may be used to construct systemic risk indexes with theoretically attractive asymptotic properties. Most importantly, we demonstrate their significant forecasting power in our empirical setting. We derive these estimators as a solution to the following statistical problem. Suppose all systemic risk measures are imperfectly measured versions of an unobservable systemic risk factor. Furthermore, suppose that the conditional quantiles of macroeconomic variables also depend on the unobserved factor. How may we identify this latent factor that drives both measured systemic risk and the distribution of future macroeconomic shocks?

The first solution is principal components quantile regression (PCQR). This two-step procedure first extracts principal components from the panel of systemic risk measures and then uses these factors in predictive quantile regressions.[4] The second solution is partial quantile regression (PQR), which is an adaptation of partial least squares to the quantile setting. We prove that both approaches consistently estimate conditional quantiles of macroeconomic shocks under mild conditions. We also show that PQR, our preferred estimator, produces consistent quantile forecasts with typically fewer factors than PCQR.[5]

A set of new stylized facts emerges from our empirical investigation. First, we find that a select few systemic risk measures possess significant predictive content for the downside quantiles of macroeconomic shocks such as innovations in IP growth or the Chicago Fed National Activity Index. Measures of financial sector equity volatility perform well in a variety of specifications; other variables, including leverage and liquidity measures, work well in some specifications but not others.

Next, we find that dimension-reduced systemic risk indexes reveal robust dependence between systemic risk and the probability of future negative macroeconomic shocks. In particular, our novel PQR estimator achieves significant forecast improvements across macroeconomic variables in a wide range of specifications.

Third, systemic risk measures are more informative about the left tail of macroeconomic shocks than about their central tendency or right tail. This is evident not only for systemic risk indexes, but is uniformly true across individual measures as well. This supports the idea that systemic risk is an inherently asymmetric and nonlinear phenomenon, a feature emphasized in much of the theoretical literature.

Next, we show that measures of financial sector equity volatility are the most useful individual predictors of macroeconomic downturns. In contrast, equity volatility in the nonfinancial sector appears to have little, if any, predictive power. This suggests that economic mechanisms connecting aggregate stock market volatility to the real economy, such as the uncertainty shocks mechanism in Bloom (2009), may blur an important distinction between

[3] See, for example, Bernanke and Gertler (1989), Kiyotaki and Moore (1997), Bernanke, Gertler, and Gilchrist (1999), Brunnermeier and Sannikov (2014), Gertler and Kiyotaki (2010), Mendoza (2010), and He and Krishnamurthy (2012).

[4] The use of principal components to aggregate information among a large number of predictor variables is well-understood for least squares forecasting—see Stock and Watson (2002) and Bai and Ng (2006). The use of principal components in quantile regression has been used by Ando and Tsay (2011).

[5] The key difference between PQR and PCQR is their method of dimension reduction. PQR condenses the cross section according to each predictor's quantile covariation with the forecast target, choosing a linear combination of predictors that is a consistent quantile forecast. On the other hand, PCQR condenses the cross section according to covariance within the predictors, disregarding how closely each predictor relates to the target. Dodge and Whittaker (2009) discuss a version of PQR but do not analyze its sampling properties.

uncertainty in the financial sector and uncertainty in other industries.

Finally, we find that systemic risk indicators forecast policy decisions. A rise in systemic risk predicts an increased probability of a large drop in the Federal Funds rate, suggesting that the Federal Reserve takes preventive action amid elevated risk levels. Combined with these predictable drops in the Fed Funds rate, our main result implies that such preventative action fails to fully counteract the risk of economic downturns that accompanies severe financial system distress.

In summary, our results reach a positive conclusion regarding the empirical systemic risk literature. When taken altogether, systemic risk measures indeed contain useful information regarding the probability of future macroeconomic downturns. This conclusion is based on out-of-sample tests and is robust across different choices of left tail quantiles, macroeconomic variables, and geographic region.

The remainder of the paper proceeds as follows. Section 2 defines and provides a quantitative description of a set of systemic risk measures in the US and Europe. In Section 3, we examine the power of these measures for predicting quantiles of macroeconomic shocks using univariate quantile regressions. In Section 4, we define PCQR and PQR estimators, discuss their properties, and use them to form predictive systemic risk indexes. Section 5 discusses stylized facts based on our empirical results. Section 6 concludes. The online appendix contains proofs and Monte Carlo evidence regarding PCQR and PQR estimators and other supporting material.

2. Data

This section outlines our construction of systemic risk measures and describes the macroeconomic outcomes that we study – data can downloaded from www.sethpruitt.net/GKPwebdata.zip. US measures are based on data for financial institutions identified by two-digit SIC codes 60 through 67 (finance, insurance, and real estate).[6] We obtain equity returns for US financial institutions from CRSP and book data from Compustat.

We also construct measures for Europe. Our "EU" measures pool data on financial institution equity returns from France, Germany, Italy, and Spain, which are the largest continental European Union economies. Our "UK" measures are for the United Kingdom. UK and EU returns data are obtained from Datastream.[7] We do not construct measures that require book data for the UK and EU, nor do we have data for some counterparts to US measures such as the default spread. When reporting forecasts of foreign macroeconomic outcomes, we use the US version of the risk measure if it is not available in the foreign data set.

Table 1
Sample start dates.
Measures begin in the stated year and are available through 2011 with the exception of Intl. spillover, which runs through 2009, and GZ, which runs through September 2010.

	US	UK	EU
Absorption	1927	1973	1973
AIM	1926	–	–
CoVaR	1927	1974	1974
ΔCoVaR	1927	1974	1974
MES	1927	1973	1973
MES-BE	1926	1973	1973
Book lvg.	1969	–	–
CatFin	1926	1973	1973
DCI	1928	1975	1975
Def. spr.	1926	–	–
ΔAbsorption	1927	1973	1973
Intl. spillover	1963	–	–
GZ	1973	–	–
Size conc.	1926	1973	1973
Mkt lvg.	1969	–	–
Volatility	1926	1973	1973
TED spr.	1984	–	–
Term spr.	1926	–	–
Turbulence	1932	1978	1978

2.1. Overview of measures

Bisias, Flood, Lo, and Valavanis (2012) categorize and collect definitions of more than 30 systemic risk measures. We construct variables from that survey to the extent that we have access to requisite data, and refer readers to the online appendix and to Bisias, Flood, Lo, and Valavanis (2012) for additional details. In addition, we study the CatFin measure of Allen, Bali, and Tang (2012) and the Gilchrist and Zakrajsek (2012) credit spread measure, which are not included in Bisias, Flood, Lo, and Valavanis (2012) but are relevant to our analysis. Below we provide a brief overview of the measures that we build, grouped by their defining features.

We are interested in capturing systemic risk stemming from the core of the financial system, and thus construct our measures using data for the 20 largest financial institutions in each region (US, UK, and EU) in each period.[8] Whenever the systemic risk measure is constructed from an aggregation of individual measures (for example, in the case of CoVaR, which is defined at the individual firm level), we compute the measure as an equal-weighted average of the 20 largest institutions. The only exception is size concentration of the financial sector for which we use the largest 100 institutions (or all institutions if they number fewer than 100). Table 1 shows the available sample period for each measure by region.

[6] This definition of financial sector corresponds to that commonly used in the literature (see, e.g., Acharya, Pedersen, Philippon, and Richardson, 2010).

[7] Datastream data require cleaning. We apply the following filters. (1) When a firm's data series ends with a string of zeros, the zeros are converted to missing, since this likely corresponds to a firm exiting the data set. (2) To ensure that we use liquid securities, we require firms to have nonzero returns for at least one-third of the days that they are in the sample, and we require at least three years of nonzero returns in total. (3) We winsorize positive returns at 100% to eliminate large outliers that are likely to be recording errors.

[8] If less than 20 institutions are available, we construct measures from all available institutions, and if data for fewer than 10 financial institutions are available the measure is treated as missing.

2.1.1. Institution-specific risk

Institution-specific measures are designed to capture an individual bank's contribution or sensitivity to economy-wide systemic risk. These measures include CoVaR and ΔCoVaR from Adrian and Brunnermeier (2011), marginal expected shortfall (MES) from Acharya, Pedersen, Philippon, and Richardson (2010), and MES-BE, a version of marginal expected shortfall proposed by Brownlees and Engle (2011).

2.1.2. Comovement and contagion

Comovement and contagion measures quantify dependence among financial institution equity returns. We construct the Absorption Ratio described by Kritzman, Li, Page, and Rigobon (2011), which measures the fraction of the financial system variance explained by the first K principal components (we use $K = 3$). We also construct the Dynamic Causality Index (DCI) from Billio, Lo, Getmansky, and Pelizzon (2012) which counts the number of significant Granger-causal relations among bank equity returns, and the International Spillover Index from Diebold and Yilmaz (2009) which measures comovement in macroeconomic variables across countries.[9]

2.1.3. Volatility and instability

To measure financial sector volatility, we construct two main variables. First, we compute the average equity volatility of the largest 20 financial institutions and take its average as our "volatility" variable. In addition, we construct a "turbulence" variable, following Kritzman and Li (2010), which considers returns' recent covariance relative to a longer-term covariance estimate.

Allen, Bali, and Tang (2012) propose CatFin as a value-at-risk (VaR) measure derived by looking at the cross section of financial firms at any one point in time. Such a VaR measure for financial firms is well-suited to provide an alternative measure of financial sector volatility.[10]

Motivated by the fact that loan ratios forecast GDP growth in crises (Schularick and Taylor, 2012), we calculate aggregate book leverage and market leverage for the largest 20 financial institutions. We also compute size concentration in the financial industry (the market equity Herfindal index), which captures potential instability in the sector.

2.1.4. Liquidity and credit

Liquidity and credit conditions in financial markets are measured by Amihud's (2002) illiquidity measure (AIM) aggregated across financial firms, the TED spread (LIBOR minus the T-bill rate), the default spread (BAA bond yield minus AAA bond yield), the Gilchrist and Zakrajsek (2012) credit spread measure (GZ), and the term spread (the slope of the Treasury yield curve).

2.1.5. Measures not covered

Due to data constraints, particularly in terms of time series length, we do not include measures of linkages between financial institutions (such as interbank loans or derivative positions), stress tests, or credit default swap spreads.

2.2. Macroeconomic data

Our analysis focuses on real macroeconomic shocks measured by innovations to IP growth in the US, UK, and EU. These data come from the Federal Reserve Board for the US and OECD for the UK and EU.[11] Our sample for the US is the entire postwar era 1946–2011. For the UK, data begin in 1978. Our EU sample begins in 1994.

In robustness checks, we consider US macroeconomic shocks measured by innovations to the Chicago Fed National Activity Index (CFNAI) and its subcomponents: production and income (PI), employment, unemployment, and hours (EUH), personal consumption and housing (PH), and sales, orders, and inventory (SOI). These data come from the Federal Reserve Bank of Chicago and are available beginning in 1967.

Our focus is on how systemic risk affects the distribution of future macroeconomic shocks. We define macro shocks as innovations to an autoregression in the underlying macroeconomic series (IP growth or CFNAI). This strips out variation in the target variable that is forecastable using its own history, following the forecasting literature such as Bai and Ng (2008b) and Stock and Watson (2012).[12] We choose the autoregressive order based on the Akaike Information Criterion (AIC) for each series—typical orders are between three and six in monthly data. We perform the autoregression (AR) estimation (including the AIC-based model selection) recursively out-of-sample.[13] Finally, we aggregate monthly shocks into a quarterly shock by summing monthly innovations to put the targets on a forecast horizon that is relevant for policy-makers. Further details are available in the online appendix.

9 We do not include the volatility connectedness measure of Diebold and Yilmaz (2014). Arsov, Canetti, Kodres, and Mitra (2013) show that this is a dominant leading indicator of financial sector stress in the recent crisis. Unfortunately, the Diebold-Yilmaz index is only available beginning in 1999 and thus does not cover a long enough time series to be included in our tests.

10 Allen, Bali, and Tang's (2012) CatFin measure is the simple average of three different approaches to estimating the financial sector's VaR in any particular month. Those authors note that the three components are highly correlated. We simply use the nonparametric version of CatFin, given the high correlation between all three measures (above 99%) noted by Allen, Bali, and Tang (2012).

11 For the EU, we use the OECD series for the 17-country Euro zone.

12 This is often referred to as "pre-whitening" in the forecasting literature. An alternative to pre-whitening is to conduct Granger causality tests that control for lags of the dependent variable. The online appendix shows that Granger causality tests, using Politis and Romano's (1994) stationary bootstrap, broadly agree with our findings based on autoregression residuals. We have also performed pre-whitening with autoregressions augmented to include lagged principal components from Stock and Watson's (2012) data. This produced minor quantitative changes to our results and does not alter any of our conclusions.

13 Using the full-sample AR estimate in out-of-sample quantile forecasts has little effect on our results, as the recursively estimated AR projection is stable after only a few years of observations.

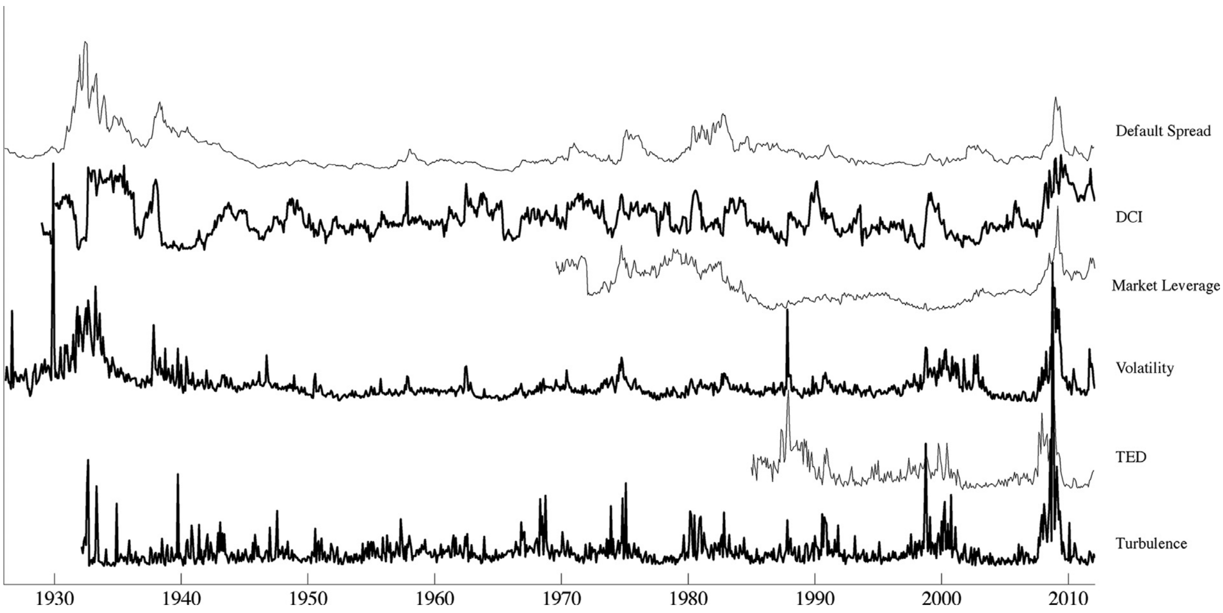

Fig. 1. Systemic risk measures. The figure plots a subset of our panel of systemic risk measures. All measures have been standardized to have equal variance.

2.3. Summary of comovement among systemic risk measures

Fig. 1 plots the monthly time series of select measures in the US.[14] All measures spiked during the recent financial crisis, which is not surprising given that many of these measures were proposed post hoc. In earlier episodes, many systemic risk measures reached similar levels to those experienced during the recent crisis. During the oil crisis and high uncertainty of the early and mid 1970s, financial sector market leverage and return turbulence spike. All the measures display substantial variability and several experience high levels in non-recessionary climates. Many of the spikes that do not seem to correspond to a financial crisis might be considered “false positives.” One interpretation of the plot is that these measures are simply noisy. Another interpretation is that these measures sometimes capture stress in the financial system that does not result in full-blown financial crises, either because policy and regulatory responses diffused the instability or the system stabilized itself (we discuss this further in Section 5.3). Yet another interpretation is that crises develop only when many systemic risk measures are simultaneously elevated, as during the recent crisis.

The online appendix reports correlations among different measures for the US, UK, and EU. Most correlations are quite low. Only small groups of measures comove strongly. For example, turbulence, volatility, and the TED spread are relatively highly correlated. Similarly, CoVaR, △CoVaR, MES, and GZ tend to comove. The other measures display low or even negative correlations with each other, suggesting that many measures capture different aspects of financial system stress or are subject to substantial noise. If low correlations are due to the former, then our tests for association between systemic risk measures and macroeconomic outcomes can help distinguish which aspects of systemic risk are most relevant from a policy standpoint.

Finally, some measures of systemic risk may be interpreted as contemporaneous stress indicators and others as leading indicators of systemic risk. We describe lead-lag relations between these variables by conducting two-way Granger causality tests in the online appendix. The GZ, default spread, turbulence, CoVaR, and volatility measures appear to behave as leading indicators in that they frequently Granger-cause other variables and not the reverse. The term spread, the international spillover index, MES, MES-BE, and DCI tend to lag other measures and thus may be viewed as coincident indicators of a systemic shock. These associations appear consistent across countries.

3. Systemic risk measures and the macroeconomy

We propose a criterion for evaluating systemic risk measures based on the relevance of each of these measures for forecasting real economic outcomes. In particular, we investigate which systemic risk measures give policy-makers significant out-of-sample information about the distribution of future bad macroeconomic shocks. We believe this criterion provides a new but natural method for evaluating policy relevance when selecting among a pool of candidate systemic risk measures.

The basic econometric tool for our analysis is predictive quantile regression, which we use to judge the relation of a systemic risk measure to future economic activity. We view quantile regression as a flexible statistical tool for investigating potentially nonlinear dynamics between systemic risk and economic outcomes. Such a reduced-form

[14] The plotted measures are standardized to have the same variance (hence no *y*-axis labels are shown) and we only show a subset of the series we study for readability.

statistical approach has benefits and limitations. Benefits include potentially less severe specification error and, most importantly, the provision of new empirical descriptions to inform future theory. A disadvantage is the inability to identify "fundamental" shocks or specific mechanisms as in a structural model. Hansen (2013) provides an insightful overview of advantages to systemic risk modeling with and without the structure of theory.

3.1. Quantile regression

Before describing our empirical results, we offer a brief overview of the econometric tools and notation that we use. Denote the target variable as y_{t+1}, a scalar real macroeconomic shock whose conditional quantiles we wish to capture with systemic risk measures. The τth quantile of y_{t+1} is its inverse probability distribution function, denoted

$$\mathbb{Q}_{\tau}(y_{t+1}) = \inf\{y : P(y_{t+1} < y) \geq \tau\}.$$

The quantile function may also be represented as the solution to an optimization problem

$$\mathbb{Q}_{\tau}(y_{t+1}) = \arg\inf_{q} E[\rho_{\tau}(y_{t+1} - q)]$$

where $\rho_{\tau}(x) = x(\tau - I_{x<0})$ is the quantile loss function.

Previous literature shows that this expectation-based quantile representation is convenient for handling conditioning information sets and deriving a plug-in M-estimator. In the seminal quantile regression specification of Koenker and Bassett (1978), the conditional quantiles of y_{t+1} are affine functions of observables $\boldsymbol{x}_t$,

$$\mathbb{Q}_{\tau}(y_{t+1}|\mathcal{I}_t) = \beta_{\tau,0} + \boldsymbol{\beta}_{\tau}'\boldsymbol{x}_t. \tag{1}$$

An advantage of quantile regression is that the coefficients $\beta_{\tau,0}$, $\boldsymbol{\beta}_{\tau}$ are allowed to differ across quantiles.[15] Thus, quantile models can provide a richer picture of the target distribution when conditioning information shifts more than just the distribution's location. As Eq. (1) suggests, we focus on quantile forecasts rather than contemporaneous regression since leading indicators are most useful from a policy and regulatory standpoint.

Our focus is on the out-of-sample information provided by systemic risk measures. In everything that follows, we are careful to construct systemic risk measures (and later on, systemic risk indexes) in a recursive out-of-sample manner. This means that the forecast of a macroeconomic shock at time $t+1$ is constructed using only information from the estimation sample $\{1, 2, \ldots, t-1, t\}$. In particular, all parameters and fitted values are estimated using data ending no later than time t.

Forecast accuracy can be evaluated via a quantile R^2 based on the loss function ρ_{τ}:

$$R^2 = 1 - \frac{\frac{1}{T}\sum_t[\rho_{\tau}(y_{t+1} - \hat{\alpha} - \hat{\beta}X_t)]}{\frac{1}{T}\sum_t[\rho_{\tau}(y_{t+1} - \hat{q}_{\tau})]}.$$

This expression captures the typical loss using conditioning information (the numerator) relative to the loss using the historical unconditional quantile estimate (the denominator). The out-of-sample R^2 can be negative if the historical unconditional quantile offers a better forecast than the conditioning variable. We arrive at a description of statistical significance for our out-of-sample estimates by comparing the sequences of quantile forecast losses based on conditioning information, $\rho_{\tau}(y_{t+1} - \hat{\alpha} - \hat{\beta}X_t)$, to the quantile loss based on the historical unconditional quantile, $\rho_{\tau}(y_{t+1} - \hat{q}_{\tau})$, following Diebold and Mariano (1995) and West (1996).[16]

Our benchmark results focus attention on the 20th percentile, or $\tau = 0.2$. This choice represents a compromise between the conceptual benefit of emphasizing extreme regions of the distribution and the efficiency cost of using too few effective observations. In the online appendix we show that results for the 10th percentile are similar. We also estimate median regressions ($\tau = 0.5$) to study systemic risk impacts on the central tendency of macroeconomic shocks.[17]

3.2. Empirical evaluation of systemic risk measures

Table 2 Panel A reports recursive out-of-sample predictive statistics. The earliest out-of-sample start dates are 1950 for the US, 1990 for the UK, and 2000 for the EU (due to the shorter data samples outside the US). We take advantage of the longer US time series to perform subsample analysis, and report results for out-of-sample start dates of 1976 and 1990.

Only financial sector volatility, CatFin, and market leverage are significant for every region and start date. Focusing on the US, Table 2 Panel A shows that book leverage, CatFin, GZ, volatility, and turbulence are significantly informative out-of-sample for all split dates. Table 3 Panel A investigates the robustness of this observation when macroeconomic shocks are measured by the CFNAI series. Since the CFNAI begins later, we consider out-of-sample performance starting in 1976. There we see that only financial sector turbulence provides significant out-of-sample predictive content for the total CFNAI index and each of its component series.[18]

Turning to the central tendency of macroeconomic shocks, Table 4 Panel A shows that systemic risk measures demonstrate substantially weaker forecast power for the median shock. The default spread, CatFin, GZ, volatility, and turbulence possess some predictive power for the median, but less than they do for lower percentiles.

[15] Chernozhukov, Fernandez-Val, and Galichon (2010) propose a monotone rearranging of quantile curve estimates using a bootstrap-like procedure to impose that they do not cross in sample. We focus attention on only the 10th, 20th, and 50th percentiles and these estimates never cross in our sample.

[16] In the online appendix, we also consider testing for the correct conditional 20th percentile coverage following Christoffersen (1998). We find somewhat similar results, in terms of accuracy and significance, for the various measures and indexes we construct using this alternative criteria, but see that the test has lower power to discriminate between risk measures in our context.

[17] We also consider upper tail ($\tau = 0.8$) quantile regressions in the online appendix to highlight the nonlinear relation between systemic risk and future macroeconomic shocks.

[18] In the online appendix, Tables A4 and A5 report that US results are broadly similar if we study the 10th rather than the 20th percentile of IP growth and CFNAI shocks.

Table 2
20th percentile IP shock forecasts.
The table reports out-of-sample quantile forecast R^2 (in percentage) relative to the historical quantile model. Statistical significance at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively; we do not test the multiple QR model. Sample is 1946–2011 for US data, 1978–2011 for UK data and 1994–2011 for EU data. Out-of-sample start date is noted for each column. Rows "Absorption" through "Turbulence" use each systemic risk measure in a univariate quantile forecast regression for IP growth rate shocks. "Multiple QR" uses all systemic risk measures jointly in a multiple quantile regression. Rows "Mean" through "PQR" use dimension reduction techniques on all the systemic risk measures. Mean is a simple average, PCQR1 and PCQR2 use one and two principal components, respectively, in the PCQR forecasting procedure, while PQR uses a single factor. "–" Indicates insufficient data for estimation in a given sample.

	US			UK	EU
Out-of-sample start:	1950	1976	1990	1990	2000
		Panel A: Individual systemic risk measures			
Absorption	−3.14	−8.86	−3.78	0.91	7.63**
AIM	2.92**	2.62	3.56*	−0.23	0.55*
CoVaR	1.37	0.86	1.79	6.83**	6.41**
ΔCoVaR	−0.79	−3.40	−0.82	6.22**	6.92*
MES	−0.46	−2.09	1.44	2.70	4.47*
MES-BE	−1.25	−1.36	−7.17	−1.10	4.65*
Book lvg.	–	2.63**	1.38***	−2.80	−3.24
CatFin	5.74***	13.27***	17.79***	6.16***	10.73***
DCI	−1.80	−1.92	−3.35	−5.18	5.63**
Def. spr.	−0.30	3.93**	8.66***	16.35***	11.70*
ΔAbsorption	−0.83	−0.06	−0.30	0.17	0.04
Intl. spillover	–	2.02*	1.01	−0.15	−1.01
GZ	–	5.26**	14.68***	−1.82	15.83**
Size conc.	−2.25	−5.93	−3.37	−3.53	−0.40
Mkt lvg.	–	10.44***	12.67***	6.81*	8.48*
Volatility	3.21**	5.62**	8.14*	7.30**	11.96***
TED spr.	–	–	9.76***	−1.01	1.10
Term spr.	0.23	2.90*	1.31	−2.64	1.27
Turbulence	3.60***	9.23***	13.01***	−3.55	−0.62
		Panel B: Systemic risk indexes			
Multiple QR	−58.18	−36.94	7.07	−32.12	−4.49
Mean	−2.26	−3.81	−11.35	−8.83	0.57
PCQR1	−0.76	1.02	1.67	8.48**	15.06***
PCQR2	2.74	7.51**	10.64**	1.47	13.73**
PQR	6.39***	13.01***	14.98***	1.42	4.32

Table 3
20th percentile CFNAI shock forecasts.
The table reports out-of-sample quantile forecast R^2 (in percentage) relative to the historical quantile model. Statistical significance at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively; we do not test the Multiple QR model. Sample is 1967–2011. Out-of-sample period starts in 1976, except for Ted Spread which begins later. Rows "Absorption" through "Turbulence" use each systemic risk measure in a univariate quantile forecast regression for the CFNAI index or sub-index in each column. "Multiple QR" uses all systemic risk measures jointly in a multiple quantile regression. Rows "PCQR1" through "PQR" use dimension reduction techniques on all the systemic risk measures. Mean is a simple average, PCQR1 and PCQR2 use one and two principal components, respectively, in the PCQR forecasting procedure, while PQR uses a single factor.

	Total	PH	PI	SOI	EUH
		Panel A: Individual systemic risk measures			
Absorption	−3.07	−1.31	−2.98	−4.17	−2.52
AIM	−4.65	−1.89	−5.33	−8.12	−3.19
CoVaR	−3.37	−0.93	−1.85	−5.91	−2.15
ΔCoVaR	−5.70	−1.12	−3.16	−5.97	−4.34
MES	−6.40	−1.87	−5.28	−8.01	−5.36
MES-BE	−2.73	−1.89	−0.59	−3.30	−3.09
Book lvg.	−2.50	−3.01	−1.48	−2.06	−2.14
CatFin	2.46	−0.62	4.78	−1.05	5.44
DCI	−2.28	0.01	−1.75	−2.20	−1.55
Def. spr.	0.69	−1.33	0.19	0.60	−0.25
ΔAbsorption	−0.58	−1.89	1.04	−0.29	0.55
Intl. spillover	−2.07	−1.27	−0.13	−2.66	−2.02
GZ	−8.23	−6.00	−4.14	−9.84	−4.83
Size conc.	−1.75	−1.12	−0.61	−4.20	−0.63
Mkt lvg.	2.61	3.56**	2.49	−0.20	3.18
Volatility	−5.26	−2.55	−2.79	−3.92	0.02
TED spr.	2.36	1.85	3.38*	2.42	−2.76
Term spr.	1.58	0.78	0.86	0.89	3.50
Turbulence	7.68**	5.26**	9.41***	7.78**	5.83*
		Panel B: Systemic risk indexes			
Multiple QR	−55.70	−72.10	−60.84	−37.01	−53.54
Mean	2.16	1.13	2.88	−0.67	−2.23
PCQR1	−6.21	−0.58	−4.93	−9.38	−2.08
PCQR2	−0.75	−0.57	−0.42	−6.09	1.90
PQR	3.68	0.45	5.27*	7.05**	4.60

In summary, we find that few systemic risk measures possess significant power to forecast downside macroeconomic quantiles. Exceptions include measures of financial sector volatility, but even these are not robust in every specification. To the extent that we find any forecasting power, it is stronger for the lower quantiles of macroeconomic shocks than for their central tendency.

4. Systemic risk indexes and the macroeconomy

Individually, many systemic risk measures lack a robust statistical association with macroeconomic downside risk. This could be because measurement noise obscures the useful content of these series, or because different measures capture different aspects of systemic risk. Is it possible, then, to combine these measures into a more informative systemic risk index?

A naive way to aggregate information across measures is to include all the systemic risk measures as multiple regressors in the same quantile regression (QR). However, multiple QR is likely to suffer from in-sample overfit due to proliferation of parameters, similar to the overfit seen in multiple regression with many predictors (see Stock and Watson, 2006).

Alternatively, we propose a statistical model in which the conditional quantiles of y_{t+1} depend on a low-dimension unobservable factor f_t, and each individual systemic risk variable is a noisy measurement of f_t. This structure embodies the potential for dimension reduction techniques to help capture information about future macroeconomic shocks present in the cross section of individual systemic risk measures. The factor structure is similar to well-known conditional mean factor models (e.g., Geweke, 1977; Sargent and Sims, 1977; Stock and Watson, 2002). The interesting feature of our model is that it links multiple observables to latent factors that drive the conditional *quantile* of the forecast target.

We present two related procedures for constructing systemic risk indexes: principal components quantile regression and partial quantile regression. We show that they consistently estimate the latent conditional quantile driven by f_t, and we verify that these asymptotic results are accurate approximations of finite sample behavior using numerical simulations. We also show that they are

Table 4
Median IP shock forecasts.
The table reports out-of-sample quantile forecast R^2 (in percentage) relative to the historical quantile model. Statistical significance at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively; we do not test the Multiple QR model. Sample is 1946–2011 for US data, 1978–2011 for UK data and 1994–2011 for EU data. Out-of-sample start date is noted for each column. Rows "Absorption" through "Turbulence" use each systemic risk measure in a univariate quantile forecast regression for IP growth rate shocks. "Multiple QR" uses all systemic risk measures jointly in a multiple quantile regression. Rows "PCQR1" through "PQR" use dimension reduction techniques on all the systemic risk measures. Mean is a simple average, PCQR1 and PCQR2 use one and two principal components, respectively, in the PCQR forecasting procedure, while PQR uses a single factor. "–" Indicates insufficient data for estimation in a given sample.

	US			UK	EU
Out-of-sample start:	1950	1976	1990	1990	2000
Panel A: Individual systemic risk measures					
Absorption	−0.92	0.99	1.62	−1.22	1.73
AIM	−0.03	−2.08	0.16*	−0.34	−0.00
CoVaR	−0.03	−4.47	0.71	−1.55	1.45
ΔCoVaR	−0.62	−4.56	−0.05	−0.84	0.58
MES	−0.57	−4.17	−0.79	−1.61	−0.35
MES-BE	−1.47	−0.50	−0.38	3.81**	0.17
Book lvg.	–	−1.87	0.42	2.02**	−1.84
CatFin	0.59	0.89	6.85***	2.55*	2.30*
DCI	−1.69	−0.80	−0.96	−1.05	0.67
Def. spr.	−0.62	3.23**	4.91***	0.66	−2.28
ΔAbsorption	−0.83	−0.62	−0.28	−0.08	−0.18
Intl. spillover	–	−1.39	−0.67	−2.15	−0.03
GZ	–	0.51	7.59**	6.24**	2.44
Size conc.	−3.42	−1.05	−3.43	−1.45	−5.25
Mkt lvg.	–	−0.26	3.20*	1.52*	−0.70
Volatility	0.73	−0.84	4.71*	3.61**	3.62**
TED spr.	–	–	2.13**	−2.47	−1.39
Term spr.	−0.02	−0.58	−0.38	−0.38	−1.71
Turbulence	1.33**	2.69*	4.45*	0.33	−0.50
Panel B: Systemic risk indexes					
Multiple QR	−32.21	−28.30	0.12	−25.08	−14.66
Mean	1.18**	3.23***	5.31***	−1.93	−1.79
PCQR1	−1.35	−5.19	4.39	0.56	−0.42
PCQR2	0.43	−5.17	2.72	0.22	−0.58
PQR	−3.11	−1.49	5.54**	−2.30	−8.21

empirically successful, demonstrating robust out-of-sample forecasting power for downside macroeconomic risk.

4.1. A latent factor model for quantiles

We assume that the τth quantile of y_{t+1}, conditional on an information set $\mathcal{I}_t$, is a linear function of an unobservable univariate factor f_t:[19]

$$\mathbb{Q}_\tau(y_{t+1}|\mathcal{I}_t) = \alpha f_t.$$

This formulation is identical to a standard quantile regression specification, except that f_t is latent. Realizations of y_{t+1} can be written as $\alpha f_t + \eta_{t+1}$ where η_{t+1} is the quantile forecast error. The cross section of predictors (systemic risk measures) is defined as the vector $\boldsymbol{x}_t$, where

$$\boldsymbol{x}_t = \boldsymbol{\Lambda} \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \equiv \boldsymbol{\phi} f_t + \boldsymbol{\Psi} \boldsymbol{g}_t + \boldsymbol{\varepsilon}_t.$$

Idiosyncratic measurement errors are denoted by $\boldsymbol{\varepsilon}_t$. We follow Kelly and Pruitt (2013; 2015) and allow $\boldsymbol{x}_t$ to depend on the vector $\boldsymbol{g}_t$, which is an additional factor that drives the risk measures but does not drive the conditional quantile of y_{t+1}.[20] Thus, common variation among the elements of $\boldsymbol{x}_t$ has a portion that depends on f_t and is therefore relevant for forecasting the conditional distribution of y_{t+1}, as well as a forecast-irrelevant portion driven by $\boldsymbol{g}_t$. For example, $\boldsymbol{g}_t$ may include stress in financial markets that never metastasizes to the real economy or that is systemically remedied by government intervention. Not only does $\boldsymbol{g}_t$ serve as a source of noise when forecasting y_{t+1}, but it is particularly troublesome because it is pervasive among predictors.

4.2. Estimators

We propose two dimension reduction approaches that consistently estimate the conditional quantiles of y_{t+1} as the numbers of predictors and time series length simultaneously become large. We first prove each estimator's consistency and then test their empirical performance.

One can view our latent factor model as being explicit about the measurement error that contaminates each predictor's reading of f_t. The econometrics literature has proposed instrumental variables solutions and bias corrections for the quantile regression errors-in-variables problem.[21] We instead exploit the large N nature of the predictor set to deal with errors-in-variables. Dimension reduction techniques aggregate large numbers of individual predictors to isolate forecast-relevant information while averaging out measurement noise.

We list requisite assumptions in the online appendix. They include restrictions on the degree of dependence between factors, idiosyncrasies, and quantile forecast errors in the factor model just outlined. They also impose regularity conditions on the quantile forecast error density and the distribution of factor loadings.

In addition to PCQR and PQR, we consider an index that equals the simple mean of the available systemic risk measures each period. This will not be a consistent estimator of a latent factor in our model, but it is a straightforward, albeit ad hoc, benchmark for comparison.

4.2.1. Principal components quantile regression (PCQR)

The first estimator is principal component quantile regression (PCQR). In this method, we extract common factors from $\boldsymbol{x}_t$ via principal components and then use them in an otherwise standard quantile regression (the algorithm is summarized in Table 5).

PCQR produces consistent quantile forecasts when both the time series dimension and the number of predictors become large, as long as we extract as many principal components (PCs) as there are elements of $\boldsymbol{F}_t = (f_t, \boldsymbol{g}_t{}')'$.

Theorem 1 (Consistency of PCQR). *Under Assumptions 1–3, the principal components quantile regression predictor of*

[19] We omit intercept terms to ease notation in the main text; our proofs and empirical implementations include them.

[20] We assume a factor normalization such that f_t is independent of $\boldsymbol{g}_t$. For simplicity, we treat f_t as scalar, but this could be relaxed.

[21] Examples of instrumental variables approaches include Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2008), and Schennach (2008). Examples of bias correction methods include He and Liang (2000), Chesher (2001), and Wei and Carroll (2009).

Table 5
Estimators.
The predictors $\boldsymbol{x}_t$ are each time-series standardized. All quantile regressions and orthogonal quantile regressions are run for quantile τ.

Principal components quantile regression (PCQR)	
Factor stage:	Estimate $\hat{\boldsymbol{F}}_t$ by $(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{x}_t$ for $\boldsymbol{\Lambda}$ the K eigenvectors associated with the K largest eigenvalues of $\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$
Predictor stage:	Time series quantile regression of y_{t+1} on a constant and $\hat{\boldsymbol{F}}_t$
Partial Quantile Regression (PQR)	
Factor stage:	1. Time series quantile regression of y_{t+1} on a constant and x_{it} to get slope estimate $\hat{\phi}_i$ 2. Cross-section covariance of x_{it} and $\hat{\phi}_i$ for each t to get factor estimate $\hat{f}_t$
Predictor stage:	Time series quantile regression of y_{t+1} on a constant and $\hat{f}_t$

$\mathbb{Q}_\tau(y_{t+1}|\mathcal{I}_t) = \boldsymbol{\alpha}'\boldsymbol{F}_t = \alpha f_t$ is given by $\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{F}}_t$, where $\hat{\boldsymbol{F}}$ represents the first K principal components of $\boldsymbol{X}'\boldsymbol{X}/(TN)$, $K = dim(f_t, \boldsymbol{g}_t)$, and $\hat{\boldsymbol{\alpha}}$ is the quantile regression coefficient on those components. For each t, the PCQR quantile forecast satisfies

$$\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{F}}_t - \alpha' f_t \xrightarrow[N,T\to\infty]{p} 0.$$

The theorem states that PCQR builds consistent forecasts for the conditional quantile of y_{t+1}. All proofs are in the online appendix. Theorem 1 is implied by Bai and Ng's (2008a) alternative arguments for extremum estimators using PCs, or could be deduced from Ando and Tsay (2011).

4.2.2. Partial quantile regression (PQR)

For simplicity, our factor model assumes that a scalar f_t comprises all information relevant for the conditional quantile of interest. But PCQR and Theorem 1 use the vector $\hat{\boldsymbol{F}}_t$ because PCQR is only consistent if the entire factor space $(f_t, \boldsymbol{g_t}')$ is estimated. This is analogous to the distinction between principal components least squares regression and partial least squares. The former produces a consistent forecast when the entire factor space is spanned, whereas the latter is consistent as long as the subspace of relevant factors is spanned (see Kelly and Pruitt, 2015).

Our second estimator is called partial quantile regression (PQR) and extends the method of partial least squares to the quantile regression setting. PQR condenses the cross section of predictors according to their quantile covariation with the forecast target, in contrast to PCQR which condenses the cross section according to covariance within the predictors. By weighting predictors based on their predictive strength, PQR chooses a linear combination that is a consistent quantile forecast.

PQR forecasts are constructed in three stages as follows (the algorithm is summarized in Table 5). In the first pass we calculate the quantile slope coefficient of y_{t+1} on each individual predictor x_i $(i = 1, \ldots, N)$ using univariate quantile regression (denote these estimates as $\hat{\gamma}_i$).[22] The second pass consists of T covariance estimates. In each period t, we calculate the cross-sectional covariance of x_{it} with i's first-stage slope estimate. This covariance estimate is denoted $\hat{f}_t$. These serve as estimates of the latent factor realizations, f_t, by forming a weighted average of individual predictors with weights determined by first-stage slopes. The third and final pass estimates a predictive quantile regression of y_{t+1} on the time series of second-stage cross section factor estimates. Denote this final stage quantile regression coefficient as $\hat{\alpha}$.

PQR uses quantile regression in the factor estimation stage. Similar to Kelly and Pruitt's (2015) argument for partial least squares, this is done to extract only the relevant information f_t from cross section $\boldsymbol{x}_t$, while omitting the irrelevant factor $\boldsymbol{g}_t$. Factor latency produces an errors-in-variables problem in the first-stage quantile regression, and the resulting bias introduces an additional layer of complexity in establishing PQR's consistency. To overcome this, we require the additional Assumption 4. This assumption includes finiteness of higher moments for the factors and measurement errors f_t, $\boldsymbol{g}_t$, and ε_{it}, and symmetric distributions for the target-irrelevant factor $\boldsymbol{g}_t$ and its loadings, $\boldsymbol{\psi}_i$. Importantly, we do not require additional assumptions on the quantile forecast error, η_{t+1}.

Theorem 2 (Consistency of PQR). *Under Assumptions 1–4, the PQR predictor of $\mathbb{Q}_\tau(y_{t+1}|\mathcal{I}_t) = \alpha f_t$ is given by $\hat{\alpha}\hat{f}_t$, where $\hat{f}_t$ is the second-stage factor estimate and $\hat{\alpha}$ is the third-stage quantile regression coefficient. For each t, the PQR quantile forecast satisfies*

$$\hat{\alpha}\hat{f}_t - \alpha f_t \xrightarrow[N,T\to\infty]{p} 0.$$

Our proof build on arguments found in White (1994), Bai (2003), Engle and Manganelli (2004), and Angrist, Chernozhukov, and Fernandez-Val (2006). Simulation evidence in the online appendix demonstrates that our PCQR and PQR consistency results are accurate approximations of finite sample behavior. In the next section, we refer to PCQR and PQR factor estimates as "systemic risk indexes" and evaluate their forecast performance versus individual systemic risk measures.

4.3. Empirical evaluation of systemic risk indexes

Table 2 shows that PQR provides positive out-of-sample performance for the lower tail of future IP growth shocks in every region and every sample split. The improvement in R^2 over the historical quantile is 1–5% in the UK and EU. In the US, the forecast improvement is 6–15%.

Fig. 2 plots fitted quantiles for the sample beginning in 1975. The thin dashed line is the in-sample historical 20th percentile. The actual shocks are plotted as black circles alongside their forecasted values based on information three months earlier (i.e., the PQR data point plotted

[22] In a preliminary step all predictors are standardized to have equal variance, as is typically done in other dimension reduction techniques such as principal components regression and partial least squares.

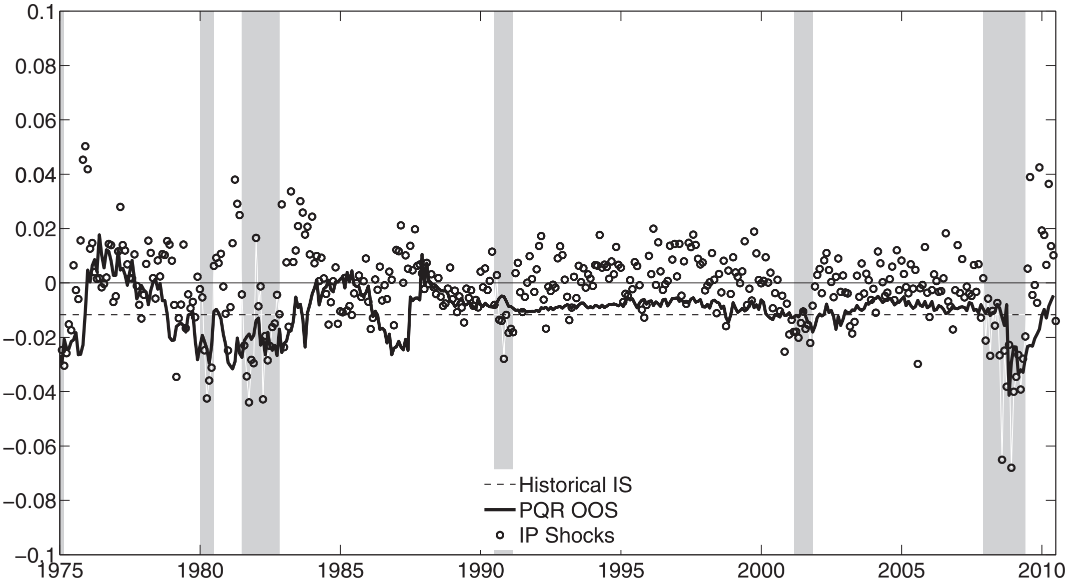

Fig. 2. IP growth shocks and predicted 20th percentiles. Fitted values for the 20th percentile of one-quarter-ahead shocks to IP growth. "Historical IS" (the thin dashed line) is the in-sample (1946–2011) 20th percentile of IP growth shocks that are shown as black circles. "PQR OOS" (the thick solid line) is the out-of-sample 20th percentile forecast based on PQR. Timing is aligned so that the one-quarter-ahead out-of-sample forecast is aligned with the realized quarterly shock. NBER recessions are shaded.

for January 2008 is the forecast constructed using information known at the end of October 2007). NBER recessions are shown in the shaded regions. The PQR-predicted conditional quantile (the thick solid line) exhibits significant variation over the last four decades, but much more so prior to the 1990s. It is interesting to note that the PQR systemic risk index predicted a large downshift in the 20th percentile of IP growth after the oil price shock of the 1970s and the recessions of the early 1980s. While the 2007–2009 financial crisis led to a downward shift in the lower quantile of IP growth, this rise in downside risk is not without historical precedent.

Table 3 Panel B shows that the PQR index also extracts positive forecasting power for the CFNAI and each subcomponent. For two of the series this forecast improvement is significant.[23]

Finally, we evaluate the ability of systemic risk indexes to forecast the central tendency of macro shocks. Table 4 Panel B shows that either PCQR or PQR rarely provide significant out-of-sample information for the median of future IP shocks.[24]

In summary, the compendium of systemic risk measures when taken together, especially in the PQR algorithm, demonstrates robust predictive power for the lower tail of macroeconomic shocks. This relation is significant when evaluated over the entire postwar period in the US, as well as in more recent sample periods in the US, UK, and EU. And while systemic risk is strongly related to lower tail risk, it appears to have little effect on the center of the distribution. This fact highlights the value of quantile regression methods, which freely allow for an asymmetric impact of systemic risk on the distribution of future macroeconomic shocks.[25]

5. Stylized facts

Our main question in this paper is whether systemic risk measures are informative about the future distribution of macroeconomic shocks. Three central facts emerge from our analysis.

5.1. Systemic risk and downside macroeconomic risk

First, systemic risk indexes are significantly related to macroeconomic lower tail risk, but not to the central tendency of macroeconomic variables. The preceding tables report significant predictability for the 20th percentile, but find little evidence of predictability for the median.

In Table 6 we formally test the hypothesis that the 20*th* percentile and median regression coefficients are equal.[26] If the difference in coefficients (20th percentile minus median) is negative, then the variable predicts a downward shift in lower tail relative to the median. Of the 22

[23] The online appendix shows that the PQR index successfully forecasts the 10*th* percentile IP growth shocks out-of-sample—the R^2 starting in 1976 is 16.5%. For the 10th percentile of CFNAI shocks, the PQR index demonstrates predictability that is statistically significant in four out of five series. The PQR forecast of the total CFNAI index achieves an R^2 of 7%.

[24] The median shock is reasonably well forecasted by the historical sample mean.

[25] We also analyze the upper tail (80th percentile forecasts) of macroeconomic shocks in the online appendix and find less out-of-sample forecasting power than for the lower tail.

[26] The *t*-statistics for differences in coefficients are calculated with a residual block bootstrap using block lengths of six months and 1,000 replications.

Table 6
Difference in coefficients, median versus 20th percentile.
In the first two columns, the table reports quarterly quantile regression coefficients for IP growth shocks at the 50th and 20th percentiles. We sign each predictor so that it is increasing in systemic risk and normalize it to have unit variance. The third column is the difference between the 20th and 50th percentile coefficients. The last column reports *t*-statistics for the difference in coefficients. Sample is the longest span for which the predictor is available.

	Median	20th pctl.	Difference	t
Absorption	−0.1936	−0.4686	−0.2750	−3.54
AIM	−0.0711	−0.0090	0.0622	0.75
CoVaR	−0.2076	−0.6946	−0.4870	−6.07
ΔCoVaR	−0.1509	−0.4963	−0.3454	−4.17
MES	−0.0980	−0.6326	−0.5346	−6.63
MES-BE	−0.0735	−0.3487	−0.2752	−3.37
Book lvg.	−0.0628	−0.1596	−0.0968	−1.20
CatFin	−0.5114	−0.7190	−0.2075	−2.65
DCI	−0.1775	−0.6132	−0.4357	−5.47
Def. spr.	−0.4237	−0.6438	−0.2202	−2.79
ΔAbsorption	0.0721	0.1110	0.0389	0.47
Intl. spillover	0.0455	−0.3459	−0.3914	−4.81
GZ	−0.5586	−0.6910	−0.1325	−1.72
Size conc.	−0.1515	−0.3256	−0.1741	−2.13
Mkt lvg.	−0.4958	−0.6243	−0.1285	−1.75
Volatility	−0.3798	−0.6675	−0.2877	−3.54
TED spr.	−0.2139	−0.5470	−0.3332	−4.14
Term spr.	0.1348	0.1372	0.0024	0.03
Turbulence	−0.5331	−0.9204	−0.3873	−4.96
Mean	−0.4119	−0.8830	−0.4710	−6.01
PCQR1	−0.4721	−0.6533	−0.1812	−2.40
PQR	−0.3086	−0.6188	−0.3102	−3.87

systemic risk measures and indexes in the table, 19 are stronger predictors of downside risk than central tendency. Of these, 16 are statistically significant at the 5% level. These results support macroeconomic models of systemic risk that feature an especially strong link between financial sector stress and the probability of a large negative shock to the real economy, as opposed to a simple downward shift in the distribution.

5.2. Financial volatility measures and economic downturns

The second stylized fact is that financial sector equity return volatility variables are the most informative individual predictors of downside macroeconomic risk.

The macroeconomic literature on uncertainty shocks, most notably Bloom (2009), argues that macroeconomic "uncertainty" (often measured by aggregate equity market volatility) is an important driver of the business cycle. Bloom shows that rises in aggregate volatility predict economic downturns.[27] Is our finding that *financial* sector volatility predicts downside macroeconomic risks merely picking up the macroeconomic uncertainty effects shown in Bloom's analysis of *aggregate* volatility? Or, instead, is the volatility of the financial sector special for understanding future macroeconomic conditions?

To explore this question, we construct two volatility variables. These are the standard deviation of daily value-weighted equity portfolio returns within each month for

Table 7
IP shock quantile forecasts: financial versus nonfinancial volatility.
The table reports out-of-sample quantile forecast R^2 (in percentage) relative to the historical quantile model. Statistical significance at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively. Sample is 1946–2011 and out-of-sample period begins in 1950. Rows use either financial or nonfinancial volatility (calculated as the average individual equity return volatility for stocks in each sector) in a quantile forecasting regression for IP growth shocks.

	80th percentile	Median	20th percentile
Financial volatility	−1.58	2.86***	5.21***
Nonfinancial volatility	−1.61	0.95	−0.72

the set of either all financial institution stocks or all nonfinancial stocks.[28] We then compare quantile forecasts of IP growth shocks based on each volatility variable.

Table 7 shows that nonfinancial volatility possesses no significant out-of-sample predictive power for the tails or median of future macroeconomic shocks. Financial volatility is a significant predictor of both central tendency and lower tail risk, but is relatively more informative about the lower tail.

These findings are consistent with the view of Schwert (1989), who uses a present value model to argue that the "rational expectations/efficient markets approach implies that time-varying stock volatility (conditional heteroskedasticity) provides important information about future macroeconomic behavior." His empirical analysis highlights comovement among aggregate market volatility, financial crises, and macroeconomic activity. Our empirical findings offer a refinement of these facts. First, they indicate that volatility of the financial sector is especially informative regarding macroeconomic outcomes compared to volatility in nonfinancial sectors. Second, they suggest that stock volatility has predictive power for macroeconomic downside outcomes (recessions) in addition to central tendency.[29]

Motivated by the result that financial volatility, and not nonfinancial volatility, provides information about the future central tendency of IP shocks, we consider an extension of Bloom's (2009) vector autoregression (VAR) analysis. We include financial volatility as an additional VAR element and study the dynamic response of IP growth.

We closely follow Bloom's (2009) original work, to make sure that any difference in results is due entirely to decomposing uncertainty into financial and nonfinancial components.[30] In particular, we estimate a nine-

[27] Recent papers such as Baker, Bloom, and Davis (2012) and Orlik and Veldkamp (2013) expand this line of research in a variety of dimensions.

[28] The volatility variable studied in preceding quantile regressions is the average equity volatility across financial firms, an aggregation approach that is consistent with our aggregation of other firm-level measures of systemic risk. The variable described here is volatility of returns on a portfolio of stocks, which is directly comparable to the market volatility variable studied in Bloom (2009).

[29] Schwert (2011) studies the association between stock volatility and unemployment in the recent crisis and notes that the extent of comovement between these two variables was weaker during the recent crisis than during the Great Depression.

[30] We follow the VAR specification shown in Bloom's (2009) Figure A1, where aggregate volatility is included in the VAR directly. IP and volatility are Hodrick-Prescott (HP) detrended using a smoothing parameter of 129,600.

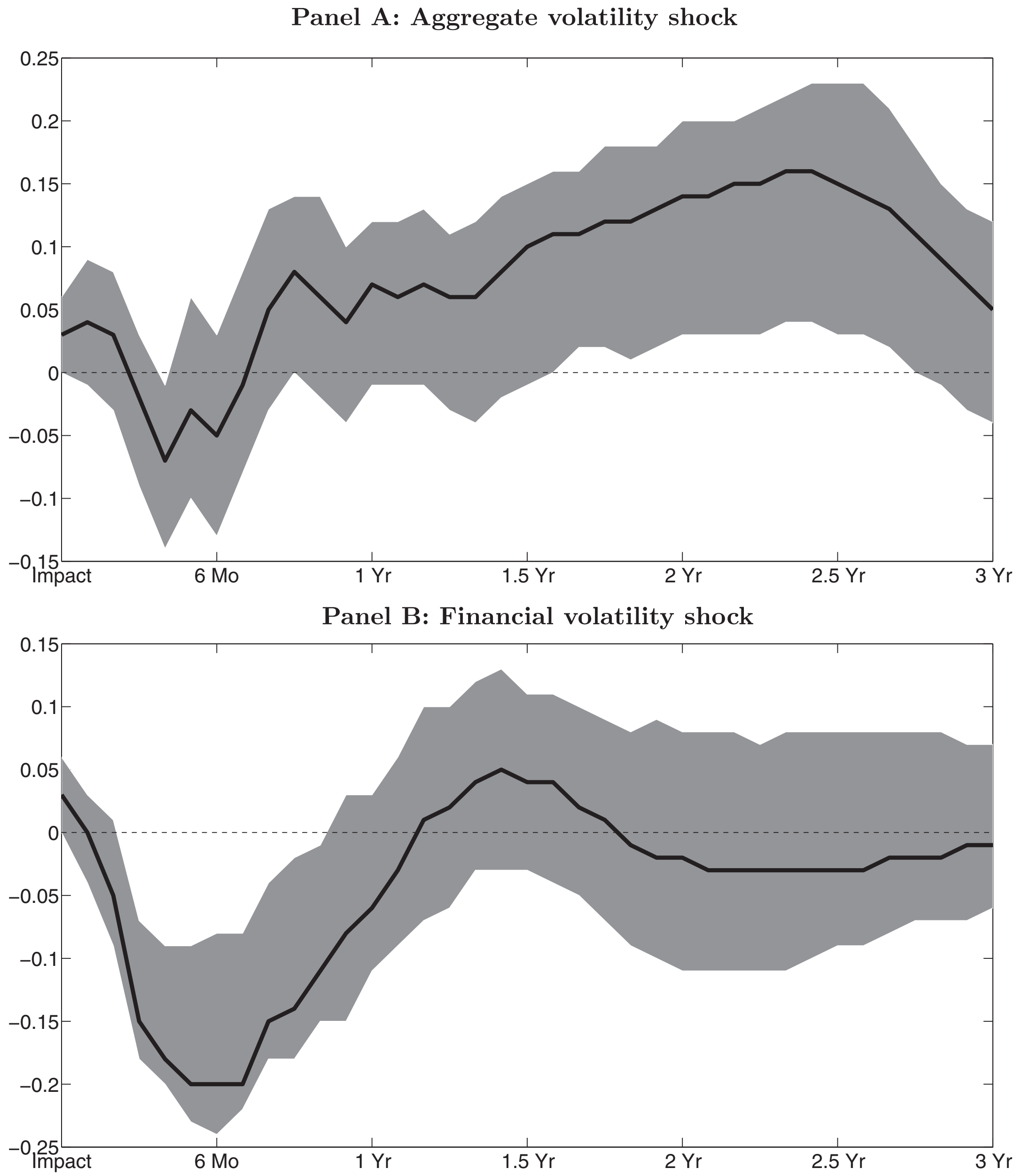

Fig. 3. Impulse response function of IP: financial volatility before aggregate volatility. Impulse response functions from Bloom's (2009) VAR, with financial volatility ordered before aggregate volatility. Orthogonal shocks identified by a Cholesky decomposition. Gray area is one-standard-error band from bootstrapping with 1,000 simulations. We use log IP, aggregate volatility, and financial volatility that have been Hodrick-Prescott (HP) detrended using smoothing parameter 129,600. Vertical axis is in log percent deviations from trend.

variable VAR, adding financial volatility to Bloom's original eight-variable specification (using exactly the same data for those eight variables, available from his website). Bloom uses a Cholesky decomposition to identify structural shocks and their effects on IP log deviations from trend. We place financial volatility either immediately before or immediately after aggregate stock market volatility, and report results in both cases.

In Fig. 3, financial volatility is ordered ahead of aggregate volatility. We plot the impulse response of IP to a financial volatility shock and to the orthogonalized aggregate volatility shock. The latter is essentially a shock to nonfinancial volatility that moves aggregate volatility but keeps financial volatility constant. As seen in the figure, we find that the financial volatility shock drives out any significant negative effect of the aggregate volatility shock.

In Fig. 4, we instead order aggregate volatility first. Here, we find that financial volatility still has a significantly negative effect on IP, even after first controlling for shocks to aggregate volatility. Note that in this version we are effectively studying a shock to the composition of volatility: The total level of volatility is held constant, but

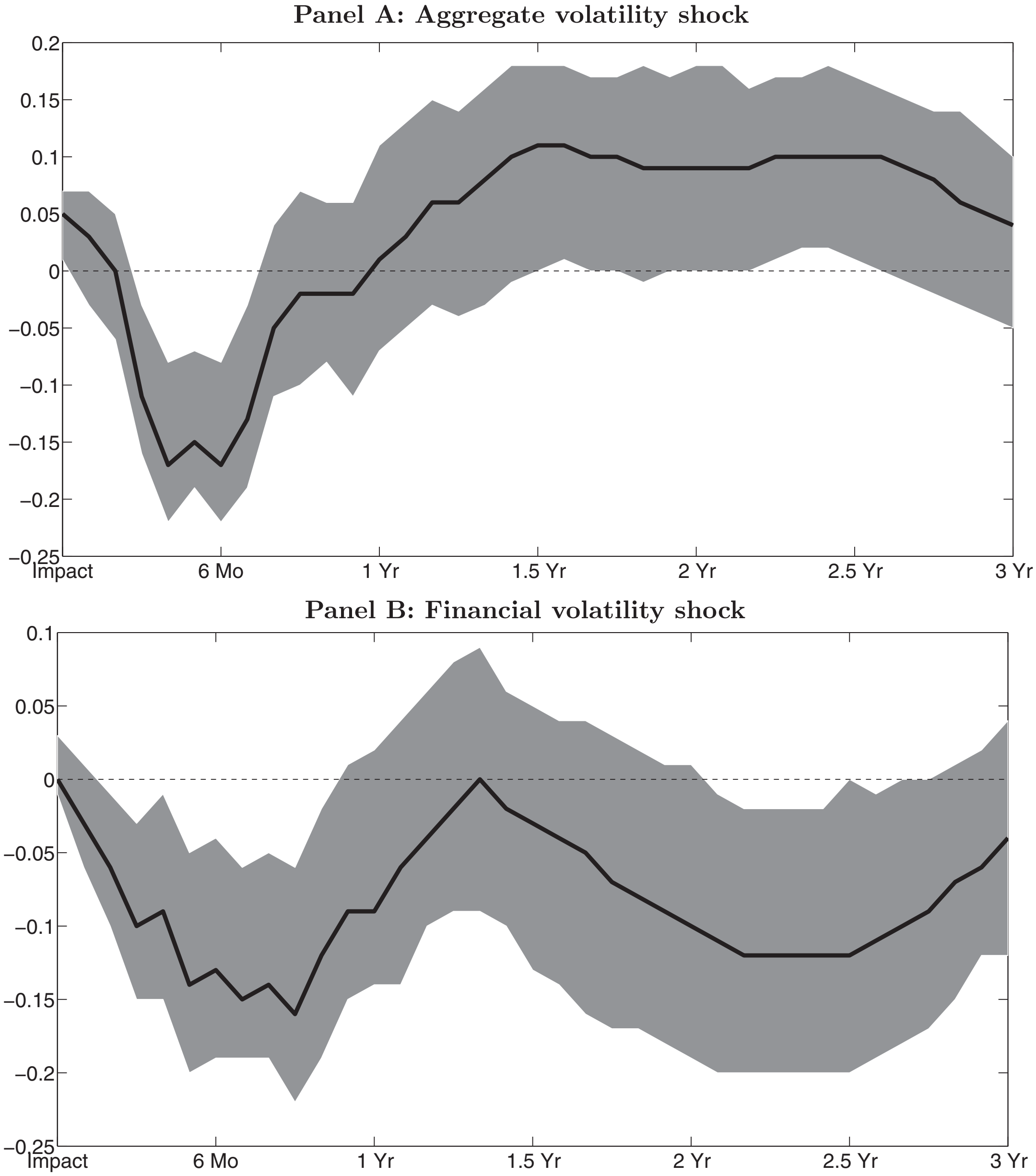

Fig. 4. Impulse response function of IP: aggregate volatility before financial volatility. Impulse response functions from Bloom's (2009) VAR, with aggregate volatility ordered before financial volatility. Orthogonal shocks identified by a Cholesky decomposition. Gray area is a one-standard-error band from bootstrapping with 1,000 simulations. We use log IP, aggregate volatility, and financial volatility that have been Hodrick-Prescott (HP) detrended using smoothing parameter 129,600. Vertical axis is in log percent deviations from trend.

the composition is shifted from nonfinancial firms toward financial firms.

In both cases, we find that the response of IP to a financial volatility shock remains negative for years afterwards. This contrasts with the shock to aggregate volatility that leads to a "volatility overshoot" where IP is above trend for 1.5–3 years after the shock.

The take-away from our predictive quantile and VAR evidence is that financial volatility plays a special role in predicting future macroeconomic activity. There are many possible explanations for why this is the case. One possibility, suggested in part by the VAR analysis, is that nonfinancial volatility can reflect good news about the future macroeconomy (as during the tech boom of the late 1990s), whereas financial volatility is predominantly bad news and reflects a weakening in the financial system's ability to efficiently match capital with projects.

5.3. Federal funds policy and systemic risk

The third stylized fact we identify is that systemic risk indicators predict an increased probability of monetary

Table 8
Federal funds rate shock forecasts.
The table reports out-of-sample quantile forecast R^2 (in percentage) relative to the historical quantile model.
Statistical significance at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively. Sample is 1960–2011. Out-of-sample begins 1965. Rows "Volatility" and "Turbulence" report univariate quantile forecast regressions on quarterly shocks to the Federal Funds rate. Row "PQR" uses a single factor estimated from all systemic risk measures. Row "Bond factors" uses three factors extracted from Fama-Bliss bond series.

	Median	20th pctl.
Volatility	1.00	3.98*
Turbulence	1.34**	3.06*
PQR	0.15	8.86***
Bond factors	−1.12	−4.01

policy easing. To show this, we examine how the Federal Reserve responds to fluctuations in various systemic risk measures. Historically, monetary policy was the primary tool at the disposal of policy-makers for regulating financial sector stress. To explore whether policy responds to systemic risk indicators, we therefore test whether the indicators predict changes in the Federal Funds rate. As in our earlier analysis, we use quantile regression to forecast the median and 20th percentile of rate changes. For brevity, we restrict our analysis to three predictor variables: financial sector volatility, turbulence, and the PQR index of all systemic risk measures.

Results reported in Table 8 show that all three measures have significant out-of-sample predictive power for the 20th percentile of rate changes. Furthermore, the out-of-sample 20th percentile predictive coefficient is significantly larger than the median coefficient, indicating that these predictors are especially powerful for forecasting large policy moves.

Is there any information in asset prices themselves that could predict these sharp movements in the Federal Funds rate? To explore this question, we test if the Treasury yield curve possesses predictive power for the quantiles of the federal funds rate. The term structure contains forward-looking information about the future path of interest rates. Thus, level, slope, and curvature of the yield curve might reflect investor beliefs regarding policy responses to the current level of systemic risk.

The last row of Table 8 reports that the yield curve does not contain predictive information about the conditional distribution of shocks to the Fed Funds rate. A potential explanation for this finding is that crises develop more rapidly than non-crisis recessions, and policy-makers scramble to respond quickly, making it difficult for investors' crisis policy expectations to show up in the standard (slow-moving) term structure factors.[31] While this test is obviously not conclusive, it highlights the usefulness of "tail risk" measures, as opposed to standard economic indicators like interest rates, for understanding the distribution of future macroeconomic shocks.

If Federal Funds rate reductions are effective in diffusing systemically risky conditions before they affect the real economy, then we would fail to detect a relation between systemic risk measures and downside macroeconomic risk. But our earlier analysis shows that the lower tail of future macroeconomic shocks shifts downward amid high systemic risk. This implies that monetary policy response is insufficient to stave off adverse macroeconomic consequences, at least in the most severe episodes.

6. Conclusion

In this paper we quantitatively examine a large collection of systemic risk measures proposed in the literature. We argue that systemic risk measures should be demonstrably associated with real macroeconomic outcomes if they are to be relied upon for regulation and policy decisions. We evaluate the importance of each candidate measure by testing its ability to predict quantiles of future macroeconomic shocks. This approach is motivated by a desire to flexibly model the way distributions of economic outcomes respond to shifts in systemic risk. We find that only a few individual measures capture shifts in macroeconomic downside risk, but none of them do so robustly across specifications.

We then propose two procedures for aggregating information in the cross section of systemic risk measures. We motivate this approach with a factor model for the conditional quantiles of macroeconomic activity. We prove that PCQR and PQR produce consistent forecasts for the true conditional quantiles of a macroeconomic target variable. Our results lead to a positive conclusion regarding the empirical systemic risk literature. When appropriately aggregated, these measures contain robust predictive power for the distribution of macroeconomic shocks.

We present three new stylized facts. First, systemic risk measures have an especially strong association with the downside risk, as opposed to central tendency, of future macroeconomic shocks. The second is that financial sector equity volatility is particularly informative about future real activity, much more so than nonfinancial volatility. The third is that financial market distress tends to precede a strong monetary policy response, though this response is insufficient to fully dispel increased downside macroeconomic risk. These empirical findings can potentially serve as guideposts for macroeconomic models of systemic risk going forward.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.jfineco.2016.01.010.

References

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70, 91–117.

[31] Because of this, it would be particularly interesting to study the prices of short maturity interest rate derivatives, such as swaptions, which would allow researchers to hone in on short-term policy expectations immediately after crisis fears begin to surface. Unfortunately, these data are traded over-the-counter, thus limited information is available, but it remains a valuable question for future research.

Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2010. Measuring systemic risk. Working paper. NYU.
Adrian, T., Brunnermeier, M., 2011. Covar. Unpublished.
Allen, L., Bali, T., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? Review of Financial Studies 25, 3000–3036.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.
Ando, T., Tsay, R., 2011. Quantile regression models with factor-augmented predictors and information criterion. Econometrics Journal 14, 1–24.
Angrist, J., Chernozhukov, V., Fernandez-Val, I., 2006. Quantile regression under misspecification, with an application to the u.s. wage structure. Econometrica 74, 539–563.
Arsov, I., Canetti, E., Kodres, L., Mitra, S., 2013. 'Near-coincident' indicators of systemic stress. Unpublished.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.
Bai, J., Ng, S., 2006. confidence intervals for diffusion index forecasts and inference for factor augmented regressions. Econometrica 74, 1133–1150.
Bai, J., Ng, S., 2008a. Extremum estimation when the predictors are estimated from large panels. Annals of Economics and Finance 9, 201–222.
Bai, J., Ng, S., 2008b. Forecasting economic time series using targeted predictors. Journal of Econometrics 146, 304–317.
Baker, S., Bloom, N., Davis, S., 2012. Measuring economic policy uncertainty. Unpublished.
Bernanke, B., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. American Economic Review 79, 14–31.
Bernanke, B., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1. Elsevier, North Holland, pp. 1341–1393.
Billio, M., Lo, A., Getmansky, M., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics 104, 535–559.
Bisias, D., Flood, M., Lo, A., Valavanis, S., 2012. A survey of systemic risk analytics. Working paper. Office of Financial Research.
Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.
Brownlees, C., Engle, R., 2011. Volatility, correlation and tails for systemic risk measurement. Unpublished.
Brunnermeier, M., Sannikov, Y., 2014. A macroeconomic model with a financial sector. American Economic Review 104, 379–421.
Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78, 1093–1125.
Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: a robust inference approach. Journal of Econometrics 142, 379–398.
Chesher, A., 2001. Quantile driven identification of structural derivatives. Unpublished.
Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.
Diebold, F., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Economic Journal 119, 158–171.
Diebold, F., Yilmaz, K., 2014. On the network topology of variance decompositions: measuring the connectedness of financial firms. Journal of Econometrics 182, 119–134.
Dodge, Y., Whittaker, J., 2009. Partial quantile regression. Biometrika 70, 35–57.
Engle, R., Manganelli, S., 2004. Caviar: conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.
Gertler, M., Kiyotaki, N., 2010. Financial intermediation and credit policy in business cycle analysis. In: Friedman, B.M., Woodford, M. (Eds.), Handbook of Monetary Economics, 3, pp. 547–599.
Geweke, J., 1977. The dynamic factor analysis of economic time series models. In: Aigner, D.J., Goldberger, A.S. (Eds.), Latent Variables in Socio-Economic Models. North Holland, Amsterdam, pp. 365–383.
Gilchrist, S., Zakrajsek, E., 2012. Credit spreads and business cycle fluctuations. American Economic Review 102, 1692–1720.
Hansen, L., 2013. Challenges in identifying and measuring systemic risk. In: Brunnermeier, M., Krishnamurthy, A. (Eds.), Risk Topography: Systemic Risk and Macro Modeling. University of Chicago Press/NBER, Chicago, IL, pp. 15–30.
He, X., Liang, H., 2000. Quantile regression estimate for a class of linear and partially linear errors-in-variables models. Statistica Sinica 10, 129–140.
He, Z., Krishnamurthy, A., 2012. A macroeconomic framework for quantifying systemic risk. Unpublished.
Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. Journal of Finance 1721–1756.
Kelly, B., Pruitt, S., 2015. The three-pass regression filter: a new approach to forecasting with many predictors. Journal of Econometrics 186, 294–316.
Kiyotaki, N., Moore, J., 1997. Credit cycles. Journal of Political Economy 105, 211–248.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Kritzman, M., Li, Y., 2010. Skulls, financial turbulence, and risk management. Financial Analysts Journal 66, 30–41.
Kritzman, M., Li, Y., Page, S., Rigobon, R., 2011. Principal components as a measure of systemic risk. Journal of Portfolio Management 37, 112–126.
Mendoza, E., 2010. Sudden stops, financial crises, and leverage. American Economic Review 100, 1941–1966.
Orlik, A., Veldkamp, L., 2013. Understanding uncertainty shocks and the role of black swans. Unpublished.
Politis, D., Romano, J., 1994. The stationary bootstrap. Journal of the American Statistial Association 89, 1303–1313.
Sargent, T., Sims, C., 1977. Business cycle modeling without pretending to have too much a priori economic theory. In: Sims, C., et al. (Eds.), New Methods of Business Cycle Research. Federal Reserve Bank of Minneapolis, Minneapolis, pp. 145–168.
Schennach, S., 2008. Quantile regression with mismeasured covariates. Econometric Theory 24, 1010–1043.
Schularick, M., Taylor, A., 2012. Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. American Economic Review 102, 1029–1061.
Schwert, W., 1989. Business Cycles, Financial Crises, and Stock Volatility. Carnegie-Rochester Conference Series on Public Policy, 31. North-Holland, pp. 83–125.
Schwert, W., 2011. Stock volatility during the recent financial crisis. European Financial Management 17, 789–805.
Stock, J., Watson, M., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.
Stock, J., Watson, M., 2006. Forecasting with many predictors. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting, 1. Elsevier, North Holland, pp. 515–554.
Stock, J., Watson, M., 2012. Generalized shrinkage methods for forecasting using many predictors. Journal of Business and Economic Statistics 30, 481–493.
Wei, Y., Carroll, R., 2009. Quantile regression with measurement error. Journal of the American Statistical Association 104, 1129–1143.
West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.
White, H., 1994. Estimation, Inference, and Specification Analysis. Cambridge University Press.