Financial conditions and density forecasts for US output and inflation

Piergiorgio Alessandri (Bank of Italy)

Haroon Mumtaz (Queen Mary, University of London)

Financial conditions and density forecasts for US output and inflation.

Piergiorgio Alessandri[*]   Haroon Mumtaz[†]

March 14, 2014

Abstract

When do financial markets help in predicting economic activity? With incomplete markets, the link between financial and real economy is state-dependent and financial indicators may turn out to be useful particularly in forecasting "tail" macroeconomic events. We examine this conjecture by studying Bayesian predictive distributions for output growth and inflation in the US between 1983 and 2012, comparing linear and nonlinear VAR models. We find that financial indicators significantly improve the accuracy of the distributions. Regime-switching models perform better than linear models thanks to their ability to capture changes in the transmission mechanism of financial shocks between good and bad times. Such models could have sent a credible advance warning ahead of the Great Recession. Furthermore, the discrepancies between models are themselves predictable, which allows the forecaster to formulate reasonable real-time guesses on which model is likely to be more accurate in the next future.

JEL classification: C53, E32, E44, G01.

Keywords: financial frictions, predictive densities, Great Recession, Threshold VAR.

---

1 Introduction

The objective of this paper is to reassess the power of financial market information in forecasting economic activity, focusing specifically on its role in predicting "tail" macroeconomic outcomes such as the Great Recession. Exploiting monthly US data for the 1973-2012 period, we estimate a set of linear and nonlinear vector autoregressions and study how the power of financial indicators in predicting output growth and inflation changes when moving (a) from a point forecasting to a distribution forecasting perspective, and (b) from linear to regime-switching models. The key message delivered by our analysis is that both dimensions are crucial. Point forecasts from linear models give a partial and in many ways distorted view of the information conveyed by financial indicators, because they do not disentangle ordinary business cycle fluctuations from the (relatively rare, but important) cases where financial shocks become the key driver of macroeconomic dynamics. Regime-switching models, on the other hand, can take into account the state-dependent nature of the nexus between finance and real economy generated by incompleteness in financial markets, and capture the changes in the transmission mechanism associated to the outbreak of such "financial crises". We show that this feature can be exploited from a density forecasting perspective, and that a regime-switching VAR would have provided a relatively accurate assessment of the likelihood of the Great Recession based on real-time financial data.

The question of whether, when and why financial markets predict economic activity has a long history in economics. The literature has delivered mixed messages so far, in the sense that no financial indicator seems to work 'too well for too long' (Stock and Watson (2003a)). However, the 2008-9 crisis and the Great Recession provided a forceful reminder that financial market disruptions can cause large macroeconomic fluctuations, which implies that financial information must, in the right circumstances, be useful in predicting those fluctuations too. In the theoretical literature on financial frictions, the emphasis is indeed shifting from the role of financial markets as an amplification mechanism for real shocks to the idea that these markets can be the source of additional shocks with important macroeconomic implications.

One way to reconcile these views is to think that the mechanisms that link financial markets to the real economy operate in a highly nonlinear, state-dependent way. If financial disturbances have implications that depend significantly on the size of the shock and on the state of the economy when the shock hits, it might be entirely possible for financial indicators to be relatively noisy and ineffective when financial markets are functioning smoothly (as suggested by most of the forecasting literature) but highly informative when the economy is on the verge of a financial crisis (as implied by the theory). The idea of a nonlinear link between financial markets and real economy is intuitive, it has received some empirical support, and it has been formalized in a number of theoretical contributions on financial frictions in general equilibrium, many of which were stimulated by the recent financial crisis (we review them below). It has not, however, been investigated in the forecasting literature, which has mostly focussed on point forecasts and linear models.

Against this background, our contribution is threefold. First, we shift the emphasis from predicting the mean path of output and inflation to forecasting their entire distribution. By studying distributions rather than point forecasts we gain in generality, providing a more complete picture of when, or in what sense, financial indicators can be valuable predictors. Second, we compare linear models (VARs) to non-linear alternatives (Threshold VARs) that can capture the time-varying, regime-dependent role of financial markets formalized by the theory. This makes our analysis more general along another dimension. More importantly, it allows us to take into account the possibility that the informational content of financial indicators is fully revealed only by the *combination* of non-linearity and predictive distributions: finance might matter mostly, or only, in off-equilibrium paths that are ignored by linear models and/or irrelevant in terms of central forecasts. Finally, alongside some widely-used tests of unconditional forecasting performance, we present an analysis of conditional forecasting performance that asks which model(s) could be expected to be more accurate at any point in time in the past based on information available until then. This kind of information is obviously equally if not more relevant from a decision maker's perspective.

Our analysis delivers three important messages. First, the presence of a financial indicator significantly improves the predictive distribution for output generated by a linear VAR. Even in a linear model, financial indicators may turn out to be more useful in predicting "tails", namely deviations of output and inflation from their expected paths, than "means", namely the expected paths themselves. Second, Threshold VARs generate noisier central forecasts than linear VARs, but they clearly outperform them in predicting distributions. In particular, these models attached a much higher *ex ante* probability to the Great Recession on the basis of real-time financial information. The advantage of threshold models stems from their ability to capture changes in both the volatility of the underlying financial shocks and the strength of their propagation mechanisms, both of which – consistently with our theoretical priors – appear to be larger in periods of financial turmoil. Third, the discrepancies between models are themselves predictable. We find that a Bayesian forecaster would have typically been in a position to formulate a reasonable real-time guess on which model was likely to be more accurate in the next future. This predictability gives place to a decision problem where the risk preferences of the forecaster take center stage. Given our data, a risk-neutral forecaster would have often chosen a linear VAR whereas a risk-averse one could have opted for a Threshold VAR, sacrificing some mean-square accuracy in order to obtain sharper advance warnings on bad tail outcomes.

The structure of the paper is the following. In Section 2 we use a simple partial equilibrium model to illustrate how financial frictions and market incompleteness can give place to non-linearities that have important implications for the conditional distribution of agents' consumption and savings decisions. Section 3 relates our work to the existing literature. Sections 4 and 5 describe respectively the data and models used in the forecasting exercise. Section 6 presents empirical evidence on the existence of finance-driven regimes in the US, and on how the transmission mechanism of (fundamental) financial shocks depends upon the regime. In Section 7 we discuss the results of our forecasting exercise, examining in detail the 2008-2009 financial crisis and the Great Recession. Section 8 concludes.

2 Predictive densities and financial frictions.

Our work is motivated by the conjecture that financial frictions have implications that pertain specifically to predictive distributions, rather than point forecasts, and that are not captured by linear models. This idea can be easily fleshed out using a simple partial equilibrium model in the spirit of Deaton (1991) and Ludvigson (1999). Consider an agent who receives a random income y_t, has access to an asset a_t which yields an exogenous return $(1+r)$, and faces a standard consumption/saving decision. The choice is subject to a constraint on his net asset position by which $a_t \geq -\theta_t y$, where $y \equiv E(y_t)$. The agent can thus save without limits but he can only borrow up to a multiple θ_t of its expected income. Following *e.g.* Kim et al. (2010) and Den Haan and De Wind (2012), we replace the occasionally binding borrowing constraint with a smooth penalty function $\mathcal{P}$ that penalizes proximity to the borrowing limit, in the sense that $\lim_{a_t \to -\theta_t y} \mathcal{P}(a_t + \theta_t y) = -\infty$.[1] The agent's utility maximization problem thus becomes:

$$\max_{(c_t,a_t)_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left(\mathcal{U}(c_t) + \mathcal{P}(a_t + \theta_t y) \right) \tag{1}$$

$$c_t + \frac{a_t}{1+r} = a_{t-1} + y_t \tag{2}$$

$$y_t = e^{z_t}, \ z_t \sim N(0, \sigma_z) \tag{3}$$

$$\theta_t = \theta(1 - \rho_\theta) + \rho_\theta \theta_{t-1} + \epsilon_t, \ \ \varepsilon_t \sim N(0, \sigma_\varepsilon) \tag{4}$$

A financial tightening is represented here by a negative ε_t shock which unexpectedly reduces the agent's borrowing capacity (θ_t) for given fundamentals (y). The

---

[1] The advantage of resorting to a penalty function to deal with an occasionally binding constraint is simplicity: the resulting model can be solved by perturbation, and the differences between linear and nonlinear solutions and between point forecasts and predictive densities can be directly linked to subsets of parameters in the policy functions. In reality, occasionally binding constraints introduce stronger forms of non-linearity – the agent's behaviour changes abruptly with the transition between unconstrained and constrained regime. Our empirical models allow for this possibility.

Table 1: Policy functions in the II order perturbation (selected coefficients; all variables are deviations from steady state values)

	ss	rc	a_{t-1}	θ_{t-1}	z_t	ϵ_t	$a_{t-1}\theta_{t-1}$	$a_{t-1}z_t$	$a_{t-1}\epsilon_t$	$\theta_{t-1}z_t$	$\theta_{t-1}\epsilon_t$
c_t	0.984	-0.002	0.264	0.058	0.263	0.116	-0.068	-0.127	-0.135	-0.067	-0.085
a_t	-0.302	0.002	0.758	-0.060	0.754	-0.119	0.069	0.130	0.139	0.070	0.088

first-order conditions for c_t and a_t are summarized by a modified Euler equation:

$$\frac{\mathcal{U}_c(c_t)}{(1+r)} - \mathcal{P}_a(a_t + \theta_t y) = \beta E_t \mathcal{U}_c(c_{t+1}) \qquad (5)$$

Equation (5) shows that an increase in today's consumption is more costly relative to the unconstrained case because it shrinks the asset buffer towards its lower bound. We adopt a simple logarithmic specification where $\mathcal{U}(c_t) = \log(c_t)$ and $\mathcal{P}(a_t + \theta_t y) = \phi \log(a_t + \theta_t y)$,and use perturbation methods to analyze the dynamics of the model around a steady state in which the agent is sufficiently impatient to hold debt, *i.e.* $a < 0$.[2,3]

Table (1) reports the policy functions for c_t and a_t derived from a second order perturbation. The first two columns show that, as expected, a second (or higher) order solution generates a "risk correction" (rc) that implies lower consumption and higher assets (*i.e.* lower debt) in steady-state relative to the first order solution (ss). The first-order terms are intuitive and we do not discuss them for brevity. Two features of the second-order terms are worth emphasizing. The a_{t-1} interaction terms imply that consumption is more sensitive to both income shocks (z_t) and financial conditions $(\theta_{t-1}, \varepsilon_t)$ when debt is high relative to equilibrium. Furthermore, the

---

[2]Using logarithms for both $\mathcal{U}$ and $\mathcal{P}$ allows us to derive the deterministic steady state of the model analytically. Provided $\theta \geq 0$ and $\phi > 0$, a necessary and sufficient condition for the agent to choose $a < 0$ in equilibrium is $\beta < (1+r)^{-1}$. We restrict our analysis to this case because financial shocks play no role if the agent is a net saver – a result akin to that reported in proposition 2 of Jermann and Quadrini (2012). Penalty functions and perturbation methods have known limitations but are appropiate in our application because the point we make is a purely qualitative one. Sensitivity analysis confirms that the instability of high order perturbations discussed by Den Haan and De Wind (2012) is not a concern.

[3]The calibration is standard, and in any case purely indicative: $\beta = 0.90$; $r = 0.03$; $\theta = 1$; $\sigma_z = 0.1$; $\sigma_\varepsilon = 0.01$; $\rho_\theta = 0.5$; $\phi = 0.05$. We also rescale y_t so that $E(y) = 1$.

$\theta_{t-1}\varepsilon_t$ interaction term in the last column implies that the impact of the financial shock on both consumption and assets is larger when financial markets are "tight", *i.e.* when θ_t is low. The reason behind the state-dependent effect of ε_t is obvious: high levels of outstanding debt (low a_{t-1}) and/or low loan-to-income ratios (low θ_{t-1}) compress the agent's borrowing capacity, thus forcing more pronounced adjustments in consumption and asset holdings in the face of any given shock. The coefficients in the last column of Table (1) imply for instance that a loan-to-income ratio of 0.5 (half the equilibrium value) increases the impact of a financial shock on c_t and a_t by roughly 30%.

The predictions obtained with a linear forecasting model would by construction ignore all interaction terms. If a_t and θ_t can be observed, the central forecast from a nonlinear model consistent with the second-order approximation could instead capture the $a_{t-1}\theta_{t-1}$ interaction. Even in this model, however, the remaining interactions would not affect the central forecast, because $E_t(x_{t+1}u_{t+1}) = E_t x_{t+1} E_t u_{t+1} = 0$ for all states $x = (a, \theta)$ and shocks $u = (z, \varepsilon)$. It is only in the predictive distribution from the non-linear model that the state-contingent implications of the shocks would fully emerge. For instance, the model would predict an increase in the variance of c_t when the observed loan-to-income ratio θ_t is low. A first, rather obvious, conclusion that can be drawn from this example is that, as long as any of the non-linearities is quantitatively significant, tests that rely exclusively on linear models may be biased towards the null that financial indicators have little predictive power. A second, more interesting one is that, if the crucial non-linearities are those involving the shocks (*e.g.* $\theta_{t-1}\varepsilon_t$), then a test that includes linear and nonlinear models but focuses exclusively on central forecasts will be subject to the same problem. The issue in this case is not that the models are inadequate in any sense, but that point forecasts ignore by construction the key channels through which the predictive power of financial data should be revealed. This can be fully uncovered only if the analysis takes into account the interaction between the nonlinear structure of the model, $\theta_{t-1}\varepsilon_t$, and the distribution of future financial shocks ε_{t+k}.

3 Literature

Our work is most directly related to the literature on the role of financial indicators in predicting output and prices. Stock and Watson (2003a) provide a survey and a comparative evaluation of a broad set of indicators, Gilchrist et al. (2009) and Gilchrist and Zakrajsek (2012) present evidence on the predictive power of corporate credit spreads, Stock and Watson (2012) and Ng and Wright (2013) revisit the issue in the light of the financial crisis and the Great Recession. These papers focus on point forecasts rather than distributions. Furthermore, the analysis relies on linear, time-invariant models. There is an open debate on how appropriate linear frameworks are given the context. Stock and Watson (2012) conclude that the Great Recession can be explained by the occurrence of large shocks in a linear dynamic factor model, whereas Sims (2012) examines VAR representations of the data finding circumstantial evidence of structural change, and Ng and Wright (2013) put the need to better investigate parameter instability at the top of the agenda in this field. We take up this suggestion, and study distributions as well as point forecasts in order to provide a more complete assessment of the implications and potential limitations associated to linear models.

Density forecasts have received relatively little attention in macroeconomics. Clements and Smith (2000) model output and unemployment in the US using a range of univariate and bivariate models, finding that nonlinear (self-exciting threshold autoregressive) specifications have no advantages in terms of point forecasting but deliver better distributions. Our results are consistent with this conclusion. Predictive densities from larger VAR models estimated on UK or US data are presented by Cogley et al. (2005), Jore et al. (2010) and Clark (2011). A general message delivered by these analyses is that accounting explicitly or implicitly (*e.g.* through rolling estimation schemes) for parameter instability, and particularly changes in volatilities, is important in order to obtain good predictive densities. None of these papers examines specifically the interaction between real economy and financial markets. We propose instead to exploit the densities to better assess the predictive power of financial indicators, pointing to financial frictions as a specific source of non-linearity

and parameter instability.[4]

There is little doubt by now that financial disturbances play an important role in causing business cycle fluctuations alongside standard (real) fundamental shocks. This conclusion is shared by Nolan and Thoenissen (2009), Kiyotaki and Moore (2012), Christiano et al. (2008), Del Negro et al. (2010). It also emerges in models that provide an explicit description of the financial intermediation process, such as Gertler and Karadi (2011) and Gertler and Kiyotaki (2010). Jermann and Quadrini (2012) and Liu et al. (2013) study credit-constrained economies with financial shocks that shift firms' borrowing constraints, in a similar way to ε_t (see Section 2), and find that these shocks account for a significant share of investment and output volatility. This result constitutes one of the key motivations behind our work. The other one comes from the debate on the nature of the nexus between financial markets and real economy.

In most DSGE analyses, including those listed above, the models are log-linearized around a steady state in which credit constraints are binding, thus ignoring the potentially nonlinear, state-dependent nature of the mechanism through which financial shocks are propagated. The financial crisis has generated new interest in capturing this aspect explicitly, and "occasionally binding" borrowing or collateral constraints of the type discussed in Section 2 have become an increasingly popular modelling device to do so. This sort of non-linearity was traditionally thought to operate at the individual level (Deaton (1991), Ludvigson (1999)). However, Huo and Rìos-Rull (2013) develop a model with heterogenous agents and idiosyncratic risk showing that a tightening in borrowing conditions can generate deep recessions even if the fraction of agents that are credit-constrained is relatively small, while Bianchi (2011), Bianchi and Mendoza (2010), Bianchi (2012) develop frameworks where financial crises are defined precisely as episodes where some form of credit or liquidity con-

---

[4]In a similar spirit, De Nicolò and Lucchetta (2012) estimate a dynamic factor model by quantile regression and examine distributions and structural impulse-responses to study the interaction between a "financial risk" indicator and real economic activity. We use higher frequency data, test linear models against nonlinear parametric alternatives that are interpretable on the basis of a theory of financial frictions, and follow a Bayesian approach that is better suited to tackling forecasting and model selection problems.

straint binds at the aggregate level. Due to the presence of missing markets, these models exhibit a highly nonlinear behavior, with kinks or even non-monotonicity in the agents' decision rules.[5] On the empirical side, evidence of structural shifts linked to changes in financial conditions is provided for instance by McCallum (1991) and Balke (2000), who document the existence of different credit regimes in US data and find the transmission of monetary policy shocks to be significantly more powerful in periods of low credit. Li and Dressler (2011) identify credit regimes as an important cause of business cycle asymmetries. Guerrieri and Iacoviello (2013) find evidence of asymmetric responses of aggregate consumption to changes in house prices in the US, and rationalize it using a model with an occasionally binding collateral constraint on housing. These analyses place no specific emphasis on the identification and estimation of structural financial shocks and do not look at forecasting issues.[6]

Our work contributes to the debate in two ways. Firstly, we provide new evidence on the existence of distinct financial regimes in the US and on how the implications of a financial shock change depending on the regime. In this way we corroborate empirically a mechanism that underpins a broad class of nonlinear general equilibrium models with incomplete financial markets, and confirm, but qualify in an important way, the general message that financial shocks "matter" for the macroeconomy. Secondly, we show how the non-linearity can be exploited from a predictive point of view. Our results suggest that cyclical changes in the transmission mechanism of financial shocks are likely to affect (and should arguably be taken into account in) any work on macroeconomic forecasting with financial variables.

---

[5]Interestingly, the models appear capable of generating crisis-like dynamics even if the underlying exogenous shocks are purely real, as in Bianchi (2011).

[6]Simple forms of non-linearity are studied in the literature on financial disruptions and recessions (e.g. Barro and Ursua, 2009), but this typically uses cross-country, low-frequency data and relies on exogenous definitions of what constitutes a crash or a recession. Our distribution-based approach is less restrictive. Furthermore, our data allows us to focus on horizons and frequencies that are more directly relevant to monetary or macroprudential policy interventions.

4 Data and forecasting methodology

We use monthly data covering the period from March 1973 to August 2012. Industrial production index (y), consumer price index (π) and the fed funds rate $(r$, an average of daily figures) are taken from the Federal Reserve Bank of St. Louis (FRED) Database. Choosing a good proxy to describe financial market conditions is not a trivial task. We use the Financial Condition Index (fci) constructed and maintained by the Chicago Fed (see Brave and Butters (2012) and references therein). Fci is a real-time indicator extracted using dynamic factor analysis from a set of over 100 series describing money, debt, equity markets and the leverage of financial intermediaries. As such, it represents to our knowledge the broadest available summary of financial conditions in the US. This has two key advantages. First, by including fci we effectively turn our (linear or nonlinear) VARs into factor models, or FAVARs, that exploit a much larger information set than they would if we used instead a "plain vanilla" financial indicator, such as a bond spread or a credit aggregate. This minimizes the possibility that the (otherwise relatively small) size of the dataset might bias the results in favour of nonlinear models – a crucial point, given our objectives. Second, the predictive power of many financial variables is known to be unstable over time, and using a broad indicator allows us to reduce the risk of obtaining results that are too heavily affected by the idiosyncratic behavior of specific variables in specific subperiods. As a robustness check, we replicate our analysis replacing fci with the Excess Bond Premium of Gilchrist and Zakrajsek (2012). The results, documented in the Annex, show that our main conclusions hold under this alternative specification.[7] We note that both indicators have been found to improve output forecasts in linear models. The key question from our perspective is whether their role changes significantly when considering nonlinear models and predictive distributions.

---

[7]Fci appears to be overall a better predictor for industrial production than the Excess Bond Premium, which further streghtens the case for focusing on it in our discussion. The question of which indicator works best, and why, is of course an interesting one, because the answer presumably depends on which frictions are more relevant. Unlike EBP, for instance, FCI takes into account household credit and captures the role of leverage in the financial intermediation sector, both of which were important in the recent financial crisis. We leave this issue for future research.

Our econometric models are described in Section 5. All models are estimated recursively over an expanding data window. Starting from the 1973.03–1983.04 sample, this gives us a set of 354 out-of-sample, "real-time" forecasts. We examine horizons of one, three, six and twelve months. Forecasts at horizons greater than one month are obtained recursively. All predictive densities are estimated using kernel methods rather than parametric approximations as in *e.g.* Clark (2011) in order to take into account any non-normal features caused by the nonlinear nature of the models. Point forecasts are calculated as the arithmetic means of the predictive densities, and evaluated in terms of Root Mean Square Errors (RMSE). To assess the accuracy of the densities we mainly rely on log-scores (LS, see Mitchell and Wallis (2011) and references therein). We also report statistics calculated on weighted log-scores that emphasize a model's accuracy in the tails of the distribution (Amisano and Giacomini (2007), see Annex for details).

RMSEs and LS are commonly used to compare the average performance of a set of models over a given period. A key issue, however, is how a decision maker would have chosen between models in real time. Even if a nonlinear model turns out to perform well on average over a sample that includes the Great Recession, one could naturally ask (i) when the evidence in favour of nonlinearity started emerging, and (ii) how strong or convincing it was when it did emerge. We investigate this issue in two complementary ways. First, we report log-predictive Bayes factors (Mitchell and Wallis (2011); Geweke and Amisano (2010)) that summarize the differences between the cumulative log-scores of the models at each date t. This allows us to establish at which points in time, or phases of the economic cycle, the models gain or loose ground relative to one another. Second, following Giacomini and White (2006), we analyze the conditional performance of the models and test whether the differences in accuracy across models can themselves be predicted using real-time information. Giacomini and White (2006) discuss the formal test procedure and show how to derive decision criteria that exploit this kind of predictability. Intuitively, given a pair of models $\{A, B\}$, the criterion is based on a regression of the difference $(LS_t^A - LS_t^B)$ on a set of time-t covariates, and suggests to pick model A if the discrepancy is

predicted to be positive at some horizon $t + k$.[8] Both Giacomini-White criteria and Bayes factors are out-of-sample statistics calculated at each point in time on the basis of time-t information and, unlike full-sample statistics, they are available and can be used to inform choices between alternative models in real time. The two indicators, however, convey a different kind of information. Bayes factors capture the overall relative performance of model A relative to model B up to time t, whereas Giacomini-White criteria capture the short-run dynamics of the discrepancy between A and B. If for instance A persistently dominates (is dominated by) B in the first (second) half of a given evaluation period $[0, ..., t]$, the Bayes factor calculated in t will suggest that the models are approximately equivalent, whereas a Giacomini-White type of criterion will suggest that B should be preferred *as of time t* because it is likely to be more accurate in $t + 1, ..., t + h$.

5   Forecasting models

5.1   Linear VAR

The benchmark model that we use is the following Bayesian VAR(13) model

$$Y_t = c + \sum_{j=1}^{P} B_j Y_{t-j} + \Omega^{1/2} e_t, e_t \tilde{} N(0,1) \tag{6}$$

where Y_t denotes the $T \times N$ data matrix of endogenous variables described below. Following Banbura et al. (2010) we introduce a natural conjugate prior for the VAR

---

[8]We refer the reader to the annex for a brief discussion of Giacomini and White (2006) and some details on our implementation of the strategy suggested in the paper.

parameters via the following dummy observations:

$$Y_{D,1} = \begin{pmatrix} \frac{diag(\gamma_1\sigma_1...\gamma_N\sigma_N)}{\tau} \\ 0_{N\times(P-1)\times N} \\ .... \\ diag\left(\sigma_1...\sigma_N\right) \\ .... \\ 0_{1\times N} \end{pmatrix}, and\ X_{D,1} = \begin{pmatrix} \frac{J_P\otimes diag(\sigma_1...\sigma_N)}{\tau} & 0_{NP\times1} \\ 0_{N\times NP} & 0_{N\times1} \\ .... & \\ 0_{1\times NP} & c \end{pmatrix} \tag{7}$$

where γ_1 to γ_N denotes the prior mean for the coefficients on the first lag, τ is the tightness of the prior on the VAR coefficients and c is the tightness of the prior on the constant terms. In our application, the prior means are chosen as the OLS estimates of the coefficients of an AR(1) regression estimated for each endogenous variable using a training sample. As is standard for US data, we set $\tau = 0.1$. The scaling factors σ_i are set using the standard deviation of the error terms from these preliminary AR(1) regressions. Finally we set $c = 1/10000$ in our implementation indicating a flat prior on the constant. We also introduce a prior on the sum of the lagged dependent variables by adding the following dummy observations:

$$Y_{D,2} = \frac{diag\left(\gamma_1\mu_1...\gamma_N\mu_N\right)}{\lambda}, \ X_{D,2} = \begin{pmatrix} \frac{(1_{1\times P})\otimes diag(\gamma_1\mu_1...\gamma_N\mu_N)}{\lambda} & 0_{N\times1} \end{pmatrix} \tag{8}$$

where μ_i denotes the sample means of the endogenous variables calculated using the training sample. As in Banbura et al. (2010), the tightness of this sum of coefficients prior is set as $\lambda = 10\tau$. Given the natural conjugate prior, the conditional posterior distributions of the VAR parameters $B = vec([c, B_1; B_2..; B_j])$ and Ω take a simple form and are defined as

$$G\left(B \setminus \Omega\right) \tilde{} N(B^*, \Omega \otimes (X^{*\prime}X^*)^{-1}) \tag{9}$$

$$G\left(\Omega \setminus B\right) \tilde{} IW(S^*, T^*), \tag{10}$$

14

where

$$B^* = (X^{*\prime}X^*)^{-1}(X^{*\prime}Y^*) \tag{11}$$

$$S^* = \left(Y^* - X^*\tilde{B}\right)'\left(Y^* - X^*\tilde{B}\right)$$

with $Y^* = [Y; Y_{D,1}; Y_{D,2}], X^* = [X; X_{D,1}; X_{D,2}]$ and $\tilde{B}$ denoting the draw of the VAR coefficients B reshaped to be conformable with X^*. T^* denotes the number of rows of Y^*.

A Gibbs sampler offers a convenient method to simulate the posterior distribution of B and Ω by drawing successively from these conditional posteriors. We employ 20,000 iterations using the last 5000 for inference. In particular, these 5000 draws are used to produce the forecast density

$$G\left(Y_{t+K} \setminus Y_t\right) = \int G\left(Y_{t+K} \setminus Y_t, \Gamma\right) \times G\left(\Gamma \setminus Y_t\right) d\Gamma$$

where $K = 1, 2, ..12$ and $\Gamma = \{B, \Omega\}$. The forecast density can be easily obtained by simulating Y_t, K periods forward using the Gibbs draws for B and Ω. Note that we use two versions of this model: the basic one (labelled $VAR^\S$) only contains our macroeconomic variables, $Y_t = \{y, r, \pi\}$, while the expanded system (labelled VAR) adds the financial indicator to the basic specification, $Y_t = \{y, r, \pi, fci\}$, in order to gauge the role played by financial information.

5.2 Threshold VAR

The Threshold VAR, or TAR, model is defined as

$$Y_t = \left[c_1 + \sum_{j=1}^{P} B_{1,j} Y_{t-j} + \Omega_1^{1/2} e_t\right] S_t + \left[c_2 + \sum_{j=1}^{P} B_{2,j} Y_{t-j} + \Omega_2^{1/2} e_t\right](1 - S_t) \tag{12}$$

where

$$S_t = 1 \iff Z_{t-d} \leq Z^* \tag{13}$$

15

The matrix of endogenous variables in the TAR model is $Y_t = \{y, r, \pi, fci\}$. The model allows for the possibility of two regimes, where the regime is determined by the level of a threshold variable Z_{t-d} relative to an unobserved threshold level Z^*. In our application, the threshold variable is assumed to be the dth lag of the financial conditions indicator, where the delay d is assumed to be an unknown parameter. The two sets of parameters $\{c_s, B_{s,j}, \Omega_s\}$, with $s = 0, 1$, can be regarded as the reduced-form counterparts of two sets of first-order conditions associated to a structural model with an occasionally binding borrowing constraint, and corresponding respectively to the states where the constraint does or does not bind. As in the linear VAR, a financial shock is implicitly added to the set of fundamental shocks that drive the dynamics of the economy. The TAR imposes fairly tight restrictions on the relation between financial conditions and transitions across regimes: in particular, the financial indicator is assumed to cause the switch across regimes in a deterministic fashion. As an alternative we also examined a Markov-Switching VAR with endogenous transition probabilities linked to fci_t. In this specification the role of financial markets is modelled more flexibly, at the cost of a heavier parameterization of the model. This model is dominated by TAR in terms of forecasting accuracy, so we omit its discussion for brevity (a formal description of model structure, estimation and main results can be found in Annex D).

As in the BVAR model above, we impose a natural conjugate prior on the VAR parameters in the two regimes. The prior tightness is set in an identical fashion to the BVAR case. We assume a normal prior for $Z^* \tilde{} N(\bar{Z}, \bar{V})$ where $\bar{Z} = 1/T \sum_{i=1}^{T} Z_t$ and $\bar{V} = 10$. Given the scale of the financial indicators used in this paper this represents a fairly loose prior. We assume a flat prior on the delay d but limit its values between 1 and 12. We employ the Gibbs sampler introduced in Chen and Lee (1995) to simulate the posterior distribution of the unknown parameters. Given an initial value for Z^* and d, the conditional posterior for the VAR parameters in the two regimes is standard and given by equations 9 and 10. Given a draw for the VAR parameters and a value for d a random walk Metropolis Hastings step can be employed to sample Z^*. We draw candidate value of Z^*_{new} from $Z^*_{new} = Z^*_{old} + \Psi^{1/2}\epsilon$, $\epsilon \tilde{} N(0, 1)$.

The acceptance probability is given by $\frac{f(Y_t \backslash Z^*_{new}, \Xi)}{f(Y_t \backslash Z^*_{old}, \Xi)}$ where $f(.)$ denotes the posterior density and Ξ represents all other parameters in the model. We choose the scaling factor Ψ to ensure that the acceptance rate remains between 20% and 40%. Chen and Lee (1995) show that the conditional posterior for d is a multinomial distribution with probability $\frac{L(Y_t \backslash d, \Xi)}{\sum_{d=1}^{12} L(Y_t \backslash d, \Xi)}$ where $L(.)$ denotes the likelihood function. We employ 20,000 iterations of the Gibbs sampler discarding the first 15,000 as burn-in. The forecast density for TVAR is defined as:

$$G\left(Y_{t+K} \backslash Y_t\right) = \int G\left(Y_{t+K} \backslash Y_t, \Gamma\right) \times G\left(\Gamma \backslash Y_t\right) d\Gamma$$

where $\Gamma = \{B_1, \Omega_1, B_2, \Omega_2, Z^*, d\}$. Given draws from the Gibbs sampler, this object can be easily computed by iterating equations 12 and 13 K periods in the future.

6 Financial regimes in the US

Before moving to forecasting, we discuss timing and structural features of the regimes identified by the threshold model. Data and estimated regimes are displayed in Figure 1. The shaded area represent the median estimate of $1 - S_t$ obtained estimating the model over the full sample. This takes a value of 1 when the financial condition indicator is above the critical threshold (see equation (13) in Section 5): for the sake of brevity we refer to this as the "crisis" regime below. The US economy enters this regime in 1974-1975 and in the early 1980s, two periods characterized by financial volatility and contractions in real output. The regime also emerges sporadically around 1987-1988. In this case the increase in fci is likely to be caused mostly by volatility in asset prices, and it is not associated to a decline in output. The regime does not occur at all in the two decades before the outbreak of the financial crisis. The last switch takes place in 2008. The crisis regime lasts roughly two years, covering the period from the first spike in fci to the end of the contraction in industrial production. All in all, the TVAR appears to isolate in a satisfactory manner sub-periods that were characterized by (some combination of) high financial volatility, tight credit markets, and weak or negative growth. This, however, does not

say much on causality. A key issue is whether the estimates support our conjecture that financial shocks play a different role in the two regimes. The analysis sketched in Section 2 suggests that these shocks should have a stronger impact on output when financial conditions are tight, namely in the crisis regime. We resort to structural impulse-response analysis to investigate the issue.

The impulse responses are calculated using monte-carlo integration as described in Koop, Pesaran and Potter (1995). In particular, the responses are based on the following definition

$$IRF_t^S = E\left(Y_{t+k}\backslash\Psi_t, Y_{t-1}^s, \mu\right) - E\left(Y_{t+k}\backslash\Psi_t, Y_{t-1}^s\right) \tag{14}$$

where Ψ_t denotes all the parameters and hyperparameters of the VAR model, k is the horizon under consideration, $S = 0, 1$ denotes the regime and μ denotes the shock. Equation 14 states that the impulse response functions are calculated as the difference between two conditional expectations. The first term in equation 14 denotes a forecast of the endogenous variables conditioned on one of the structural shocks μ. The second term is the baseline forecast, $i.e.$ conditioned on the scenario where the shock equals zero. As described in Koop, Pesaran and Potter (1995) these conditional expectations can be approximated via a stochastic simulation of the VAR model. Note that we condition the responses on observations in each regime. For example the impulse response for regime 0 is calculated for all possible starting values in that regime Y_{t-1}^0 and the average response conditioned on this regime is obtained. To identify the shocks we adopt a simple recursive scheme where y_t, π_t and r_t appear in this order, reflecting as customary the relative sluggishness of output and prices in responding to exogenous disturbances, and fci_t is ordered last. This assumption is consistent with financial variables moving quickly in response to any news on the macroeconomic outlook. It is also conservative from our perspective, as it minimizes the risk of overestimating the role played by genuine financial shocks in explaining the dynamics of the system.

Figure 2 shows the response to a one standard deviation increase in fci. The

dynamics are qualitatively similar in the two regimes, and resemble those generated by a recessionary demand shock, with a contraction in output and (to a lesser extent) inflation, and a fall in the policy rate. The difference between regimes is stark: the drop in output is both deeper and more abrupt in a crisis, with a trough of -3% percent 3-6 months after the shock. Interestingly, the response of the policy rate is also stronger in the second regime. This does not necessarily imply that the Fed paid more attention to financial markets in 'bad times'; the result is indeed also consistent with the presence of a unique, time-invariant Taylor rule combined with larger expected responses by output and inflation in crisis periods. Since the TVAR allows the residual covariance matrix to change across regimes, these differences are partly due to the different size of the shock. The bottom panels of Figure 2 show that the standard deviation of the financial shock is indeed roughly three times larger in a crisis (0.3 versus 0.1). In order to isolate the role played by the transmission mechanism, we replicate the analysis simulating an fci shock of the same absolute size in the two regimes. We pick an increase of 0.1 units, roughly the equivalent of one standard deviation in 'good times'. As Figure 3 shows, following this shock output falls by roughly twice as much if the economy is going through a financial crisis. The transmission mechanism clearly plays an important role. Interestingly, this amplification comes about despite a much lower persistence of fci in the crisis regime.

Since the switch is endogenous, the fans might in principle include cases where the economy moves across regimes along the way, thus mixing two different transmission mechanisms. Given that the mean of the posterior for the threshold Z^* is around 0.25, though, a one standard deviation (or 0.1 units) shock is sufficiently small to make the transition from regime 1 to regime 2 extremely unlikely, so the left-hand side plots in Figures 2 and 2 can be taken to represent a genuine "good times" response. Increasing the size of the shock makes the responses more similar across regimes, because with large shocks the 'crisis' parameterization tends to dominate independently of the initial state of the economy. We find, however, that the responses differ even with sizable shocks. After a 0.5 unit increase in fci, which in normal times corresponds to five standard deviations, for instance, output falls by 5% in the crisis regime and

19

3.8% in normal times, and the decline is again far more gradual in the latter case.

We stress that, from the point of view of density prediction, the changes in volatilities and transmission mechanisms revealed by this analysis work in a complementary way. If financial shocks are both larger *and* more powerful as an output driver in 'bad times', the predictive power of fci should indeed be much higher in those periods, or when the economy is close to the threshold and hence reasonably likely to cross it over the forecasting horizon.

7   Forecast analysis

7.1   Full sample results

Table 2 shows average RMSEs and LS based on the full set of forecasts. In the table, $VAR^\S$ is the benchmark three-variable vector autoregression without financial indicator, VAR is the model that includes the Financial Condition Index, and TAR is the threshold model where the index drives the transition across regimes.[9] Note first that the presence of fci_t in the linear VAR improves the forecasts for industrial production. The decline in RMSEs observed in Tables 2 is consistent with the evidence offered by the existing literature. We find, however, that the improvement brought about by the indicator is both quantitatively larger and more robust if judged in terms of LS: by this metric, VAR dominates $VAR^\S$ at all horizons.[10] This suggests that even in a linear model financial indicators can be more useful in predicting tail outcomes, *i.e.* deviations of output from its expected path, than means, *i.e.* the expected paths themselves. Note that, on the other hand, VAR does not enjoy any clear advantage over $VAR^\S$ in predicting inflation or interest rates in a root-mean square error sense. The instability of correlations of this kind has been examined

---

[9]Standard specification (or "calibration") diagnostics such as Probability Integral Transforms (PITs), inverse PITs and Probability Coverage Ratios appear to be broadly similar across models and specifications, and are thus not particularly informative given the focus of the paper.(see Annex; more detailed results are available upon request).

[10]The improvement in root mean square errors and log-scores also emerges in the specifications where fci_t is replaced by the Excess Bond Premium of Gilchrist and Zakrajsek (2012) (see Annex).

elsewhere (e.g by Stock and Watson (2003a)), and these results are not central to our analysis, so we do not investigate them further here.

A second, crucial result is that RMSE and LS rank linear and nonlinear models in a very different way. A glance at the left-hand panels of the table shows that linear VARs (with or without financial indicators) tend to generate lower RMSEs than the TAR. 2 shows that the two linear VARs jointly generate the best forecasts for both output and inflation at all horizons. The result appears to be robust with respect to both the choice of the financial indicator and the specification of the nonlinear model. If judged by its root-mean square errors, the VAR of Table 2 is indeed in an absolute sense the best forecasting model for industrial production: it outperforms not only its direct TAR competitor, as shown in the table, but also TARs and MSVARs estimated using either the Financial Condition Index or the Excess Bond Premium of Gilchrist and Zakrajsek (2012) (see Annex C and D). The picture is completely different in terms of log-scores. As the right-hand panels of the tables show, the top spot here is taken by TAR for all specifications, variables and horizons. Broadly speaking, multiple regimes represent a liability for central forecasts and an asset for predictive densities. The densities clearly bring out features of the relation between financial variables and macroeconomic outcomes that cannot be pinned down if either the non linearity or the distributional angle are missing. The discussion in Section 2 provides a natural way to rationalize this finding.

Two more facts are worth emphasizing. First, TAR is also unambiguously better than VAR at predicting the financial indicator itself in a log-score sense (this is also true for the specifications based on the Excess Bond Premium, see Annex). This suggests that regime switches are an important intrinsic feature of the financial data we examine. Second, the TAR in Table 2 generates the most accurate distributions for output: like the corresponding VAR in RMSE space, this model outperforms not only its direct competitors, as shown in the table, but also threshold and Markov-switching models estimated using the Excess Bond Premium. Fci thus appears to be the best financial indicator in our set independently of which models (linear/nonlinear) and evaluation criteria (RSME/LS) one wishes to empha-

size. This result is not surprising given the way the indicator is constructed (see Section 4).

Table 3 provides a gauge on the statistical significance of the differences between models. The table reports pvalues on the pairwise tests of equal unconditional and conditional predictive ability discussed in Section 4. The pvalues are purely indicative because our estimates are obtained recursively. Significance naturally varies substantially across variables and models. The upshot, though, is that in most cases the difference between models would be deemed to be significant, particularly in conditional terms. That means that, even in cases where two competing models have a broadly similar performance over the full sample, their relative accuracy varies widely over specific subsamples and, importantly, does so in a somewhat predictable way – we return to this issue in the next section.

Table 4 reports weighted log-scores based on the two weighting schemes discussed in Section 4, focusing on one-month ahead predictions. As we noted above, the weighting schemes are designed to emphasize accuracy in the tails of the distributions, downweighting instead good predictions obtained around the mean of the target variable. TAR can be seen to systematically outperform the two VARs by this metric too: it is the best model for all variables under both weighting schemes. The (indicative) pvalues are relatively large in the case of y. One reason for this is the similar performance across models in the early 1990s and early 2000 recessions, during which the link between output and financial conditions was more tenuous, or more ambiguous, than in the Great Recession. This issue is discussed in the next section.

7.2   Selecting a model in real time

In order to shed light on "which model works when" in our sample, in Figures 4 to 7 we plot the 12-month ahead predictive log-scores of $VAR^{\S}, VAR$ and TAR for, respectively, y_t, π_t, r_t and fci_t. The latter are in levels, whereas for output and inflation we look at cumulative growth rates. Recessions appear to be much harder

22

to predict, even in a probabilistic sense: LS drops for all models when y_t turns negative in the early 1990s, around 2001 and most noticeably in 2009. Figure 4 allows a visual breakdown of the discrepancy between models during the Great Recession. VAR performs much better than $VAR^\S$, which misses the recession altogether, providing a clear illustration of the positive effect of introducing fci in a linear model. The model that attaches the highest *ex ante* probability to the observed output contraction is TAR. The difference between VAR and TAR appears small in the chart but is quantitatively significant – we discuss it in greater detail in Section 7.3. For inflation and the fed funds rate (Figures 5 and 6), TAR is more accurate for most of the sample period. An interesting pattern is that this model does not anticipate the sharp interest rate cuts of the early 1990s, early 2000s and 2007, but predicts more accurately the period of low, stable rates that follow these interventions. Figure 7 shows that the predictions for fci generated by the TAR are consistently more accurate in both good and bad times. Relative to the VAR, the model underestimates the likelihood of the initial rise in financial distress in early 2007, but produces a better forecast for both the peak of the crisis and the subsequent normalization in market conditions.[11]

Figure 4 shows that VAR and TAR are no better than $VAR^\S$ in predicting the recessions of 1990 and 2001. In the case of the early 1990s the result is not surprising. This recession arguably did not originate in the financial sector: its proximate cause was rather a fall in consumption due to political uncertainty linked to the invasion of Kuwait, combined with rising oil prices[12]. The 2001 recession

---

[11]As a further robustness check, we examine forecasts generated by a linear VAR under a rolling estimation scheme based on a 10-year data window instead of a recursive scheme. This provides a simple way to capture general forms of variation in the parameters that are potentially unrelated to changes in financial conditions. The rolling scheme delivers better point forecasts for r and f but worse forecasts for y and π. The log-scores are not significantly affected by the change in the estimation. Importantly, the linear VAR confirms to be less accurate than the TAR for all variables and horizons, consistent with the results displayed in Table (2). We take this as further evidence that the form of nonlinearity captured by the TAR, namely changes in financial regimes, is empirically relevant.

[12]Neither the average credit spread on nonfinancial corporate bonds constructed by Gilchrist and Zakrajsek (2012) nor its "excess bond premium" component (see Section 4) rose in or ahead of the early 1990s recession. The same was true of the spreads between AAA and BAA-rated bonds and

was mainly due to a contraction in business investment triggered by a revision in expectations on information technology, and it entailed a sharp drop in stock market quotations and tightening of credit conditions. As Stock and Watson (2003b) show, the forecasting performance of both financial and non-financial leading indicators around that episode was very heterogeneous. On the financial side, equity prices and term spreads anticipated the slowdown (and so did the Excess Bond Premium of Gilchrist and Zakrajsek (2012)) while commercial paper-bill spreads and money growth failed to do so. fci includes all of the above, as well as various measures of credit quantity and quality – such as private and public debt issuances, mortgage delinquencies, repo volumes – that did not move much ahead of the recession. As Figure 1 shows, the indicator remained indeed close to its sample mean throughout the early 2000s. By this (very broad) metric, financial conditions did not change abnormally either before or after the stock market crash, so there was no reason to expect output to fall.[13]

In order to investigate to what extent shifts in the relative performance of the models could have been exploited in real time, we now turn to pairwise model comparisons based on Bayes factors and the Giacomini-White (GW) decision criteria discussed in Section 4. We focus throughout on 12-month ahead predictions; the evaluation criteria are qualitatively similar, but more volatile, at shorter horizons. Figures 10 and 11 show GW criteria for output growth and inflation forecasts. For each variable we compare VAR and TAR in terms of both RMSE and LS. The criteria are defined in such a way that in all cases positive values indicate that the TAR is expected to perform better.[14] Figure 10 shows a clear tension between root mean square error and log-score. The VAR is consistently selected on the basis of RMSE.

---

between commercial paper and Treasury bill notes.

[13]We do not generate forecasts for the early 1980s because we use the first 10 years of data as a training sample. Our conjecture, based on the evidence in Section 6, is that the TAR could predict the "twin recession" quite accurately, as both troughs in industrial production happen at a time when the US economy is estimated to be in the credit-constrained regime.

[14]In other words, the RMSE criterion (C_t^{RMSE}) is calculated using $\Delta L_m \equiv RMSE^{VAR} - RMSE^{TAR}$, while the LS criterion (C_t^{LS}) is calculated using $\Delta L_m \equiv LS^{TAR} - LS^{VAR}$, so that both criteria are positive (negative) when the TAR is more (less) accurate.

The LS criterion, on the other hand, hovers around zero for most of the sample, but sends a clear, timely and persistent signal in favour of the TAR from the very beginning of the crisis. The gap between LS and RMSE is also apparent in the case of inflation (Figure 11), but here TAR is consistently selected as the best model for density prediction since the early 1990s.

We now turn to log-predictive Bayes factors. For each variable and pair of models, the time-t factor is calculated as the cumulative difference in log-scores among models up to time t (a definition is provided in the Annex). This provides an indication of how the evidence in favour of a given model against its competitors evolves over time. Figure 8 reports the factors for the marginal densities of each of the four variables. The top-left panel confirms that, as far as output is concerned, the models have similar performances up to 2007, but the Great Recession dramatically increases the evidence in favour of both (a) VAR over $VAR^{\S}$ *and* (b) TAR over VAR. For the remaining three variables, the evidence supporting the TAR against the two linear alternatives builds up consistently over time. The interest rate chart (top-right panel) displays swings that are consistent with the relative worsening in the performance of the model in the wake of large rate cuts (see Figure 6).

Accuracy criteria based on univariate densities place no weight on correlations, and might thus be of limited value to a forecaster wishing to assess the joint distribution of a set of target variables. At a minimum, a central bank would need information on the joint distribution of output and inflation. Figure 9 shows Bayes factors calculated on the joint predictive density of these two variables, again at the 12-month horizon. By this metric, $VAR^{\S}$ dominates in the earlier period and effectively outperforms VAR until 2007, presumably because of its accuracy in predicting inflation. Once again, TAR gains ground steadily from the early 1990s, establishes itself as the best forecasting model by 1999, and passes the Great Recession test much better than both linear VARs. Our key conclusion is thus confirmed, and possibly reinforced, once the correlation between output and inflation is taken into account.

7.3 The 2008 financial crisis

The sequence of predictions generated by the models between 2007 and 2009 (which we do not report for brevity, but make available upon request) show that none of the models foresaw the first contractions in industrial production. The gap between $VAR^\S$ and VAR or TAR, however, widens rapidly. In the case of VAR and TAR, the median forecasts for output growth up to one year ahead turn negative from around March 2008 and worsen in the following months. From August 2008 onwards, the data lie mostly well within the fans, and often not far from the median. TAR appears to outperform VAR on two accounts. First, its median prediction is generally closer to the actual outcomes. Second, the fans are generally wider, which gives a more realistic picture of the uncertainty surrounding the central forecast (and implies that prediction failures are less heavily penalized by any logscore-based evaluation criteria). $VAR^\S$ responds more slowly to the data, underestimates the persistence of the fall in output, and wrongly predicts positive or near-zero growth from the third quarter of 2008. Interestingly, this model also completely misses the fall in the fed funds rate, whereas VAR and TAR anticipate a fall in the interest rate, possibly as a consequence of their (relatively accurate) predictions on the contraction in industrial production. Inflation appears harder to forecast: the fans miss most of the negative out-turns for π_t, and have altogether a remarkably similar shape across models.

Moving beyond a purely statistical analysis of the forecasts, an important issue is how to extract from these distributions indicators that can guide actual decision making. A clear definition of the decision maker's problem and loss function is obviously crucial in this respect. In the last five years, central banks around the world have been grappling with the challenge of designing and setting up new "macroprudential" policy regimes. While the nature and the specific objectives of the underlying toolkits are still being debated, it is clear that one of their main aims will consist of controlling the likelihood and real impact of a financial bust. Having at hand a predictive distribution, a macroprudential policy maker would thus presumably ask what is the probability of a tail event – a large output loss, however defined – occurring in the next future, and act when the probability exceeds a predefined

threshold. In order to check the models potential to be used as "early warning" systems within a macroprudential framework, we thus look at the model-implied recession probabilities. Figures 12 shows the probabilities attached by $VAR^{\S}$, VAR and TAR to a year-on-year output contraction of 10% or more. That is, the event of interest is assumed to be $\Sigma_{h=1}^{12} y_{t+h} < -10\%$, where y_t denotes as usual monthly growth in industrial production, and the assessment is assumed to take place at time t. The figure shows that the probabilities fluctuate in the early part of the evaluation sample, particularly for the nonlinear model, before dropping to zero in the 1990s. In the case of $VAR^{\S}$ the probability stays at zero throughout the Great Recession. VAR and TAR on the other hand issue significant warnings, with probabilities that exceed 50% at the peak, and appear to be remarkably similar. In Figure 13 the plot is replicated taking as reference point an overall output contraction of 20%. Interesting differences emerge here between VAR and TAR. The threshold model can be seen to be marginally more timely and far more extreme in its assessment of the downside risks to the real economy compared to the linear model: it estimates a peak probability of roughly 30%, whereas the linear model remains below 10%. These statistics confirm the obvious truth that it would have been impossible to obtain useful warnings from a model without financial indicators at all, such as $VAR^{\S}$. They also show that employing a nonlinear model might turn out to be crucial in order to receive a warning that is "loud enough" for a policy maker to take action.

8   Conclusions

We re-examine the predictive power of financial indicators for real economic activity, focusing on the usefulness of financial information in forecasting "tail" macroeconomic outcomes such as the Great Recession. Our analysis places the emphasis on predictive distributions, rather than point forecasts, and takes into account the nonlinear nature of the mechanisms that link financial markets and real economy. We argue that this combination of a distributional angle and a nonlinear modelling approach can reveal aspects of the comovements between financial variables and macroeconomic aggregates that are necessarily ignored in a linear point forecast-

27

ing set up. This argument is illustrated using a simple partial equilibrium model where an occasionally binding borrowing constraint introduces a non-linearity that has important implications for the conditional distribution of agents' consumption and savings decisions. We then compare the forecasting performance of a set of linear and nonlinear (threshold) VARs estimated on a monthly US dataset covering the 1972-2012 period, and ask to what extent, and in what sense, the presence of a financial distress index can refine the models' predictions for industrial production growth and consumer price inflation.

The analysis delivers three important results. The first one is that the financial indicator significantly improves the predictive distribution for output generated by a linear VAR. This suggests that even in a linear model financial information may be more useful in predicting "tails", namely deviations of output and inflation from their expected paths, than "means", *i.e.* the expected paths themselves. We regard this result as interesting because most of the empirical literature on the predictive power of financial variables has instead focused on the latter.

The second one is that nonlinear models generate noisier central forecasts than VARs, but clearly outperform them in predicting the (marginal and joint) distribution(s) of output and inflation. In particular, a Threshold VAR would have attached a much higher *ex ante* probability to the Great Recession on the basis of real-time financial information, providing policy makers with a stronger warning on the likelihood and severity of the upcoming downturn. The advantage of the threshold model stems from its ability to capture changes in both the volatility of the underlying fundamental financial shocks and the strength of their propagation mechanisms, both of which, consistently with our theoretical priors, appear to be higher in periods of financial turmoil.

The third result is that most of the discrepancies between models are themselves predictable to some extent. We find that, historically, a Bayesian decision maker would have often been able to formulate a reasonable real-time guess on which model was likely to be more accurate in the next future. This predictability gives place to a model selection problem where the risk preferences of the forecaster take center

stage. Equipped with our data and models, a risk-neutral forecaster would have often chosen a linear VAR, whereas a risk-averse one could have opted for a Threshold VAR, sacrificing some mean-square accuracy in order to obtain sharper advance warnings on bad tail outcomes. We view this result as particularly interesting in the light of the current debate on systemic risk and macroprudential regulation. Our evidence suggests that an authority tasked with both a monetary and a macroprudential policy objective would have to think hard about the set of forecasting models it employs and the metrics by which these are assessed, and possibly use different tools for different policy purposes.

References

Amisano, G. and G. Fagan (2010, June). Money growth and inflation: a regime switching approach. Working Paper Series 1207, European Central Bank.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics 25*(2), 177–190.

Balke, N. S. (2000). Credit and economic activity: credit regimes and nonlinear propagation of shocks. *The Review of Economics and Statistics 82*(2), 344–9.

Banbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *The American Economic Review 101*(7), 3400–26.

Bianchi, J. (2012). Efficient bail-outs? mimeo.

Bianchi, J. and E. G. Mendoza (2010). Overborrowing, financial crises and macro-prudential taxes. NBER Working Paper n.16091.

Brave, S. and A. R. Butters (2012). Diagnosing the financial system: financial conditions and financial stress. *International Journal of Central Banking 8*(2), 191–239.

Chen, C. W. S. and J. C. Lee (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis 16*(5), 483–492.

Christiano, L. J., R. Motto, and M. Rostagno (2008). Financial factors in business cycles. mimeo.

Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics 29*(3), 327–41.

Clements, M. P. and J. Smith (2000). Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting 19*, 255–76.

Cogley, T., S. Morozov, and T. Sargent (2005). Bayesian fan charts for uk inflation: forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control 29*(11), 1891–1925.

De Nicolò, G. and M. Lucchetta (2012). Systemic real and financial risks: measurement, forecasting, and stress testing. IMF Working Paper 12/58.

Deaton, A. (1991). Saving and liquidity constraints. *Econometrica 59*(5), 1221–48.

Del Negro, M., G. Eggertson, A. Ferrero, and N. Kiyotaki (2010). The great escape? a quantitative evaluation of the fedŠs non-standard policies. mimeo.

Den Haan, W. J. and J. De Wind (2012). Nonlinear and stable perturbation-based approximations. *Journal of Economic Dynamics and Control 36*, 1477–97.

Filardo, A. J. and S. F. Gordon (1998, July). Business cycle durations. *Journal of Econometrics 85*(1), 99–123.

Gertler, M. and P. Karadi (2011). A model of unconventional monetary policy. *Journal of Monetary Economics 58*(1), 17–34.

Gertler, M. and N. Kiyotaki (2010). Financial intermediation and credit policy in business cycle analysis. *Handbook of Monetary Economics 3 11*, 547–99.

Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*, 216–30.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–78.

Gilchrist, S., V. Yankov, and E. Zakrajsek (2009). Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets. *Journal of Monetary Economics 56*, 471–93.

Gilchrist, S. and E. Zakrajsek (2012). Credit spreads and business cycle fluctuations. *The American Economic Review 102*(4), 1692–720.

Guerrieri, L. and M. Iacoviello (2013). Collateral constraints and macroeconomic asymmetries. International Finance Discussion Paper.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–84.

Huo, Z. and J. V. Rìos-Rull (2013). Balance sheet recessions. mimeo.

Jermann, U. and V. Quadrini (2012). Macroeconomic effects of financial shocks. *The American Economic Review 102*(1), 238–71.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining density forecasts from vars with uncertain instabilities. *Journal of Applied Econometrics 25*, 621–34.

Kim, C.-J. and C. R. Nelson (1999). *State-space models with regime switching*. Cambridge, Massachusetts: MIT Press.

Kim, S. H., R. Kollmann, and J. Kim (2010). Solving the incomplete market model with aggregate uncertainty using a perturbation method. *Journal of Economic Dynamics and Control 34*, 50–58.

Kiyotaki, N. and J. Moore (2012). Liquidity, business cycles and monetary policy. NBER Working Paper n.17934.

Li, S. M. and S. Dressler (2011). Business cycle asymmetry via occasionally binding international borrowing constraints. *Journal of Macroeconomics 33*, 33–41.

Liu, Z., P. Wang, and T. Zha (2013). Land-price dynamics and macroeconomic fluctuations. *Econometrica 81*(3), 1147–84.

Ludvigson, S. (1999). Consumption and credit: a model of time-varying liquidity constraints. *The Review of Economics and Statistics 81*(3), 434–47.

McCallum, J. (1991). Credit rationing and the monetary transmission mechanism. *The American Economic Review 81*(4), 946–51.

Mitchell, J. and K. F. Wallis (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics 26*, 1023–40.

Ng, S. and J. H. Wright (2013). Facts and challenges from the great recession for forecasting and macroeconomic modeling. *Journal of Economic Literature 51*(4), 1120–54.

Nolan, C. and C. Thoenissen (2009). Financial shocks and the us business cycle. *Journal of Monetary Economics 56*, 596–604.

Sims, C. A. (2012). Disentangling the channels of the 2007-2009 recession: a comment. Brookings Papers on Economic Activity.

Stock, J. H. and M. H. Watson (2003a). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

Stock, J. H. and M. W. Watson (2003b). How did leading indicator forecasts perform during the 2001 recession? *Federal Reserve of Richmond Economic Quarterly 89*(3), 71–90.

Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007-2009 recession. NBER Working Paper n.18094.

Table 2: Point and density prediction statistics. RMSE (LS) is the average root mean square error (log-score). y, r, π, f are industrial production growth, fed funds rate, consumer price inflation and the Financial Condition Index (see Section 4 for definitions and sources). VAR§ is a three-variable VAR without financial variables. Stars identify the best model for each criterior, variable and horizon.

		RMSE				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
VAR§	y	5.604	6.465	6.804	7.019	-3.674	-3.338	-3.418	-3.948
	r	0.167*	0.357	0.598	0.985	-0.675	-1.380	-1.754	-2.118
	π	2.078	2.607*	2.812*	3.077*	-2.584	-2.658	-2.266	-2.137
	f	–	–	–	–	–	–	–	–
VAR	y	5.446*	6.166*	6.558*	6.912*	-3.553	-3.156	-3.032	-2.964
	r	0.177	0.365	0.602	0.989	-0.645	-1.357	-1.723	-2.101
	π	2.067*	2.620	2.839	3.115	-2.583	-2.550	-2.339	-2.171
	f	0.102*	0.197	0.289	0.386	0.135	-0.649	-0.957	-1.130
TAR	y	5.491	6.187	6.594	6.934	-3.491*	-3.152*	-3.005*	-2.885*
	r	0.167	0.338*	0.555*	0.943*	0.022*	-0.778*	-1.364*	-1.999*
	π	2.115	2.667	2.864	3.116	-2.503*	-2.415*	-2.195*	-2.080*
	f	0.104	0.190	0.271	0.367	0.496*	-0.122*	-0.431*	-0.717*

Table 3: Tests of equal predictive ability. Entries are pvalues for the null hypothesis of equal pairwise unconditional (top panel) and conditional (bottom panel) accuracy. See Section 3 for details.

		RMSE				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
					Unconditional test:				
VAR§,VAR:	y	0.223	0.100	0.235	0.409	0.115	0.050	0.151	0.262
	r	0.130	0.665	0.906	0.932	0.000	0.107	0.083	0.428
	π	0.495	0.527	0.350	0.389	0.989	0.221	0.256	0.205
	f	–	–	–	–	–	–	–	–
TAR, VAR:	y	0.375	0.704	0.517	0.748	0.375	0.807	0.426	0.203
	r	0.131	0.126	0.134	0.331	0.000	0.000	0.000	0.451
	π	0.012	0.075	0.454	0.985	0.441	0.293	0.084	0.000
	f	0.542	0.437	0.276	0.297	0.038	0.000	0.000	0.000
					Conditional test:				
VAR§,VAR:	y	0.443	0.049	0.119	0.215	0.091	0.142	0.334	0.533
	r	0.126	0.004	0.002	0.011	0.000	0.002	0.021	0.166
	π	0.073	0.073	0.035	0.084	0.230	0.381	0.520	0.204
	f	–	–	–	–	–	–	–	–
TAR, VAR:	y	0.106	0.000	0.003	0.036	0.429	0.120	0.433	0.446
	r	0.174	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	π	0.050	0.064	0.007	0.000	0.739	0.484	0.043	0.000
	f	0.306	0.000	0.000	0.004	0.118	0.000	0.000	0.000

Table 4: Weighted log-scores. Average scores (top panel) and pair-wise tests of equal conditional predictive ability (bottom panel) based on the weighting schemes sugested by Amisano and Giacomini (2007). See Section 4 and footnote 5 for details. The statistics are calculated on 1-month ahead predictions.

	Left tail				Both tails			
	y	r	π	f	y	r	π	f
				Weighted log-scores:				
VAR§	-1.882	-0.513	-1.846	–	-0.924	-0.220	-0.914	–
VAR	-1.761	-0.491	-1.848	0.249	-0.816	-0.211	-0.927	-0.076
TAR	-1.698	0.037	-1.754	0.458	-0.753	0.031	-0.838	0.149
				P-values:				
VAR§,VAR	0.050	0.000	0.230	–	0.139	0.021	0.181	–
TAR, VAR	0.367	0.000	0.429	0.000	0.399	0.000	0.575	0.000
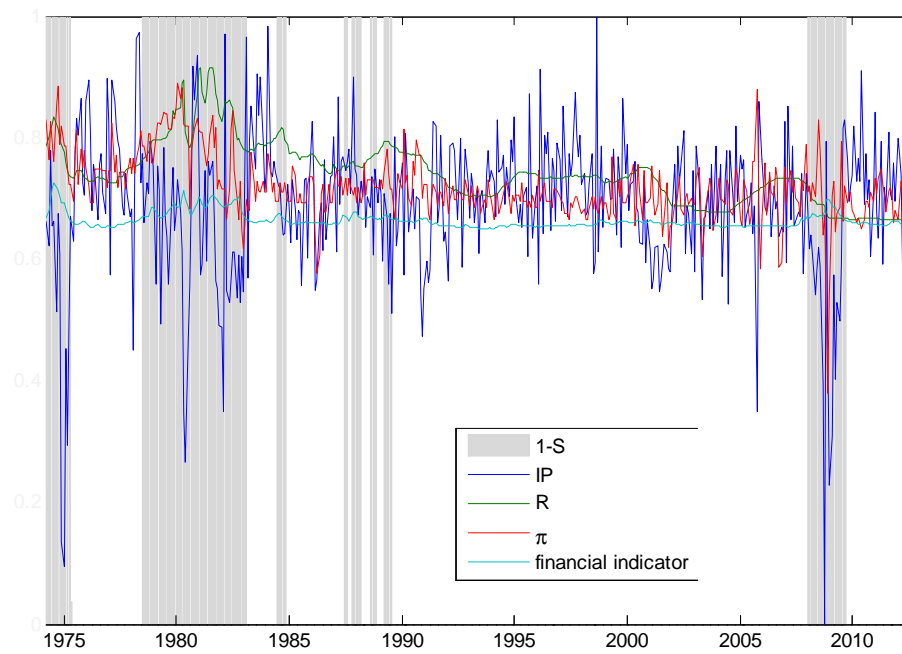
Figure 1: Financial regimes. The data is plotted against the full-sample median estimate of $1 - S_t$, the state variable that drives the regimes in the TAR model. Grey bands identify the periods of financial distress ($S_t = 0$, see equation 13).
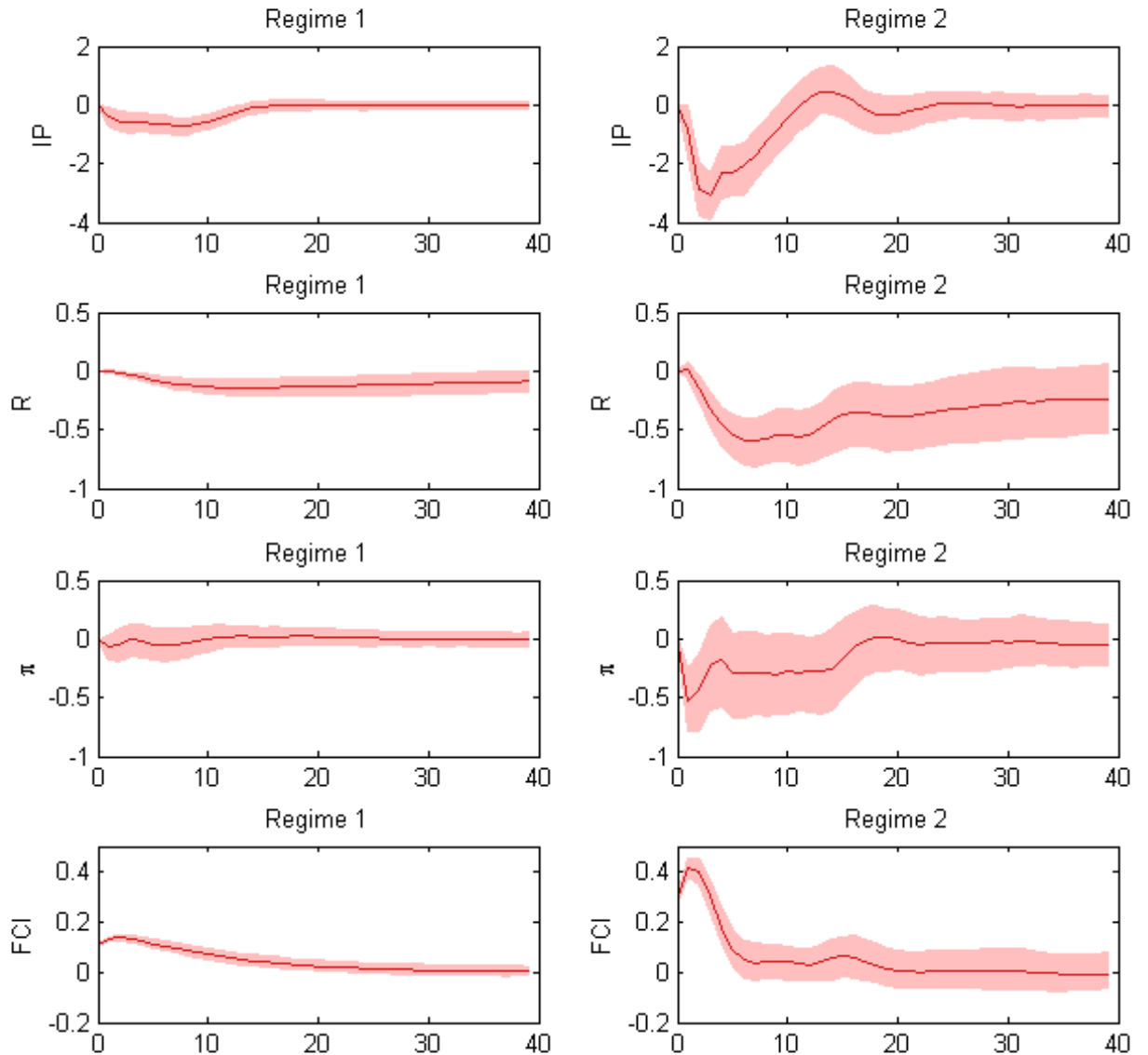
Figure 2: Impact of a one standard deviation financial shock in the TAR model. The shocks are identified recursively, ordering the financial indicator last. A positive shock implies an increase in financial distress, as measured by the Financial Condition Index (bottom panels). Regimes 1 and 2 correspond to normal times ($S = 1$) and crises ($S = 0$; see Section 6 for details).

38
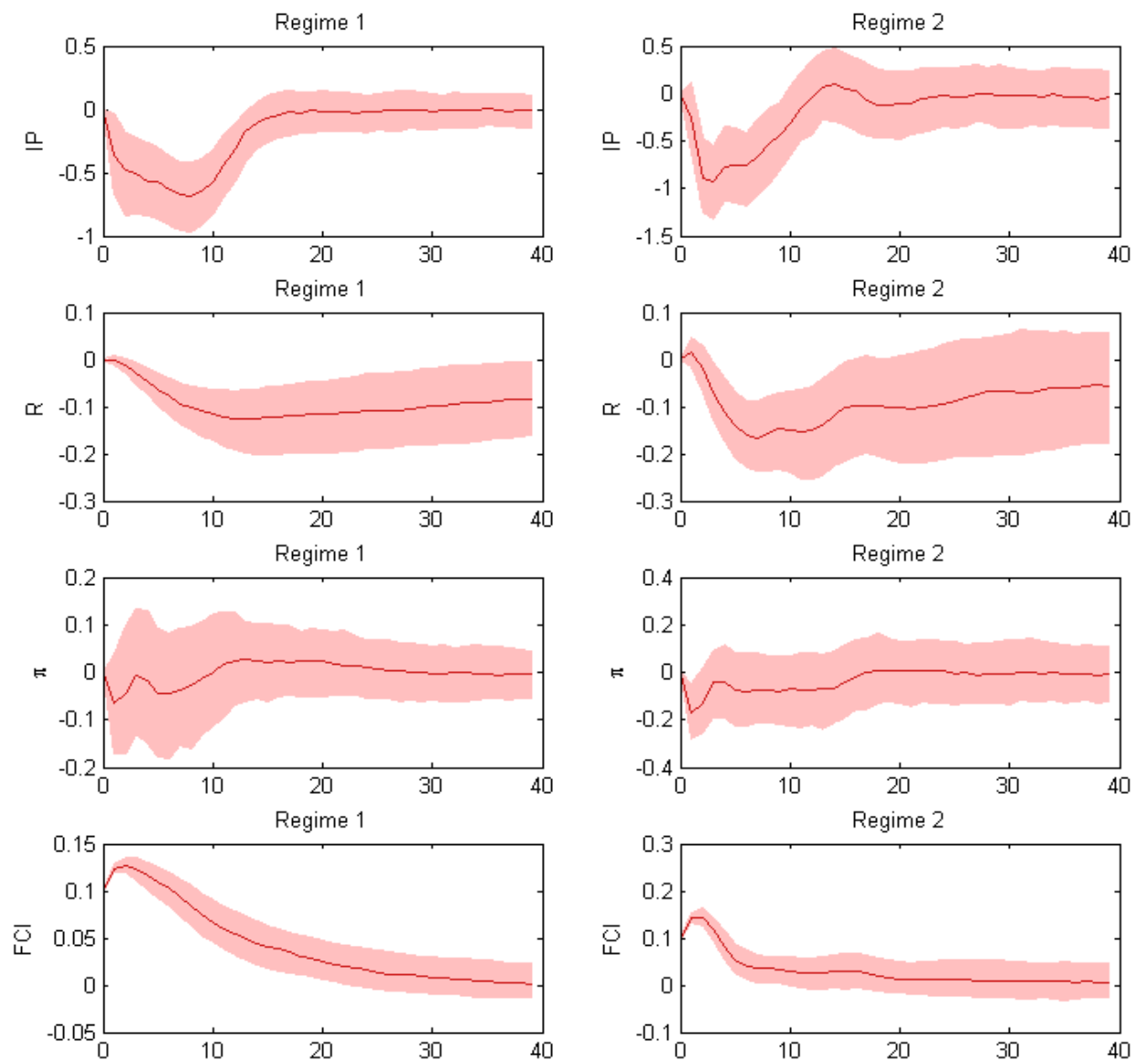
Figure 3: Response to a 0.1 unit financial shock. See note to Figure 2.
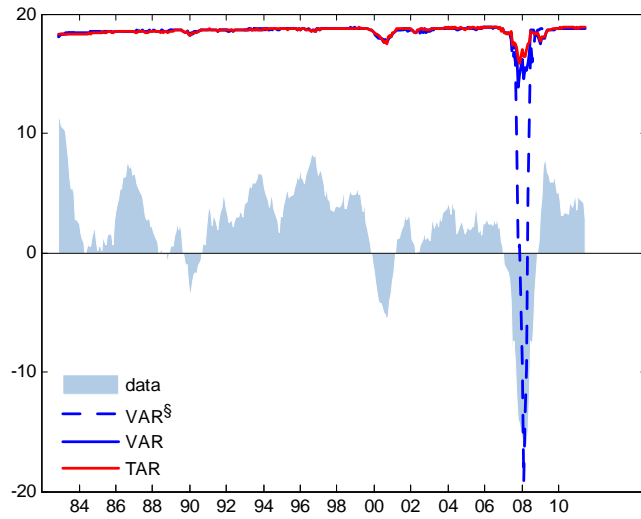
Figure 4: Log-scores, industrial production. The area shows annual growth in industrial production (left axis). The lines plot the log-score associated to the corresponding 12-month ahead predictions generated by $VAR^{\S}$, VAR and TAR (right axis).
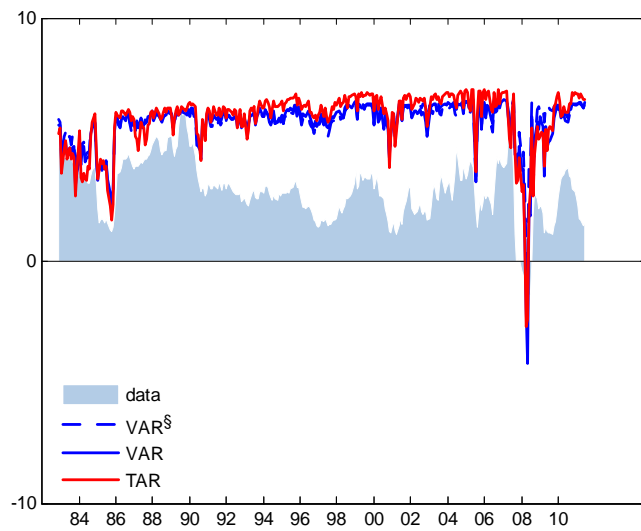
Figure 5: Log-scores, consumer price inflation. See note to Figure 4.
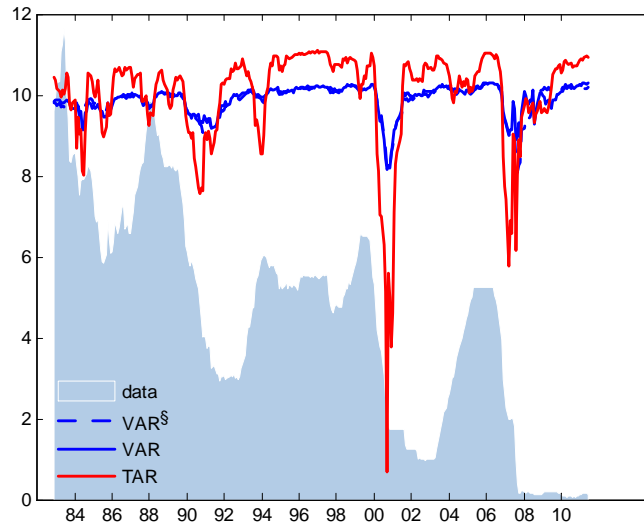
Figure 6: Log-scores, Fed funds rate. See note to Figure 4. The rate is modelled in levels.
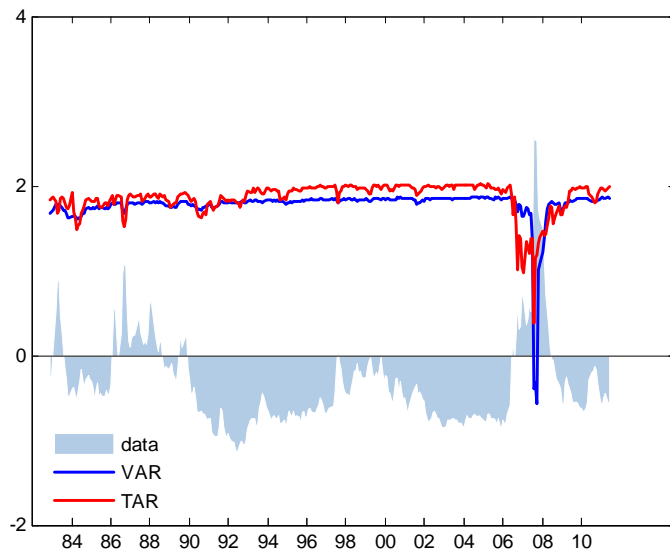
Figure 7: Log-scores, Financial Condition Index. See note to Figure 4. The index is modelled in levels.
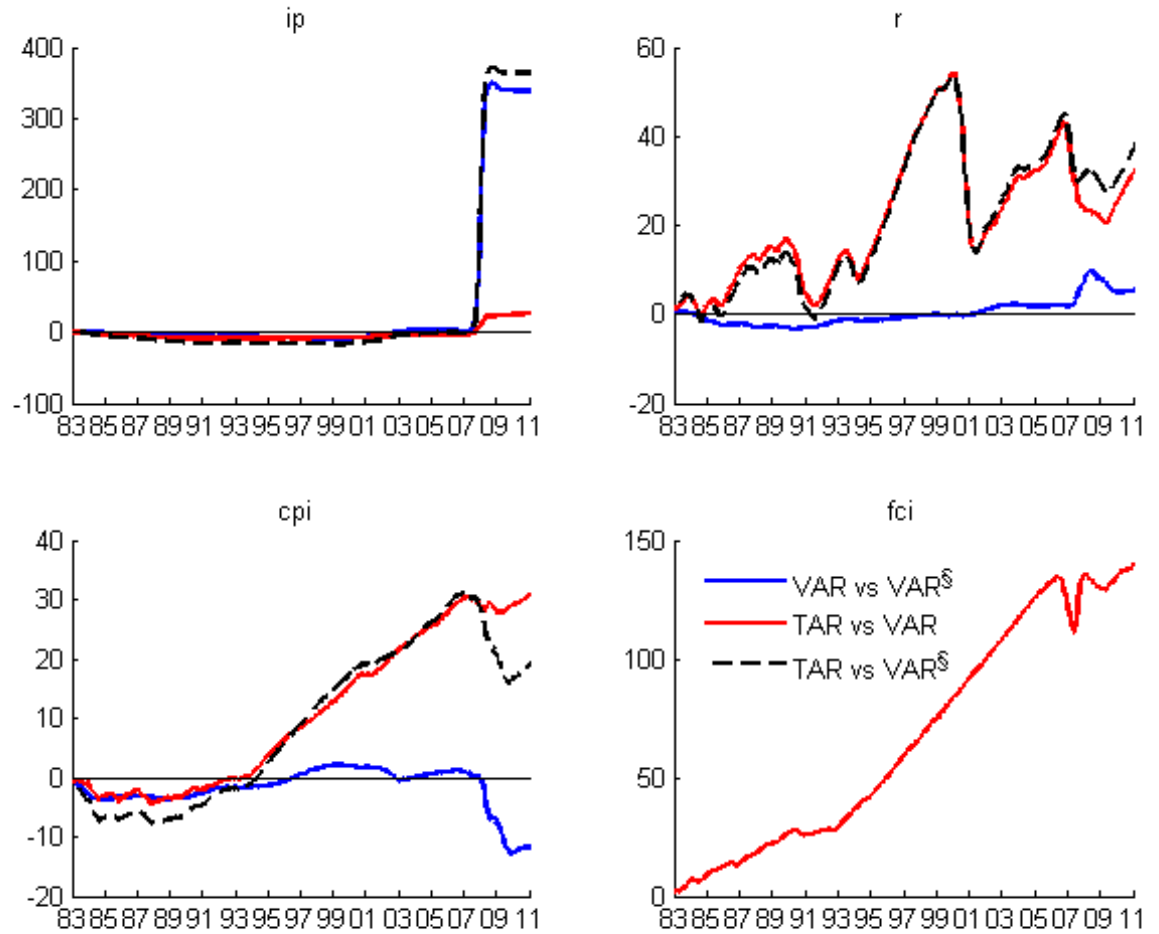
Figure 8: Log Bayes factors, marginal distributions. For each variable and each pair of models, the time-t factor is calculated as the cumulative difference in log-scores up to time t, based on 12-month ahead predictions. Ip, r, cpi, fci stand for industrial production growth, fed funds rate, consumer price inflation, and Financial Condition Index.
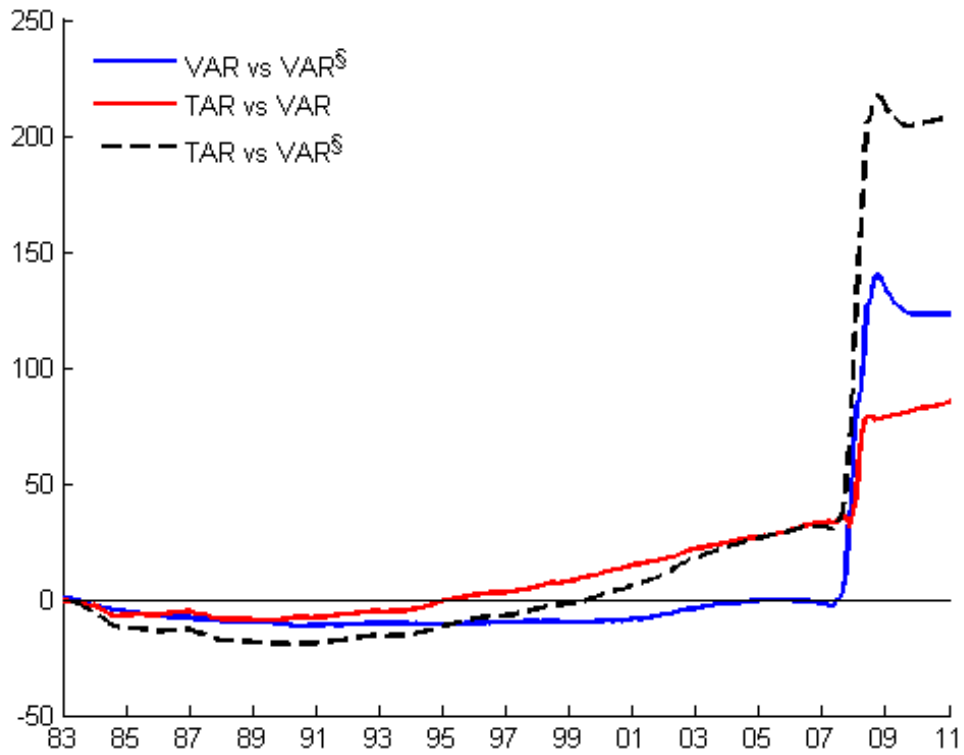
Figure 9: Log Bayes factors, joint distribution of output and inflation. See note to Figure 8. Here, the log-scores are calculated on the joint predictive density for industrial production and inflation.
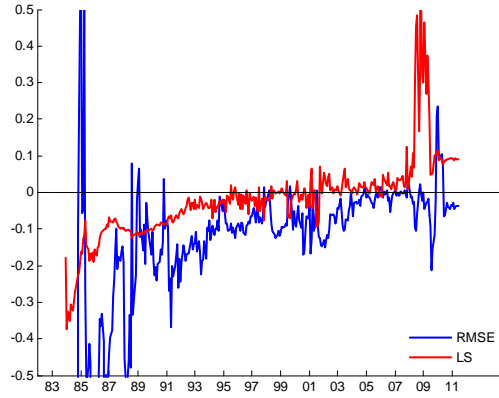
Figure 10: Real-time model selection criteria, industrial production. The criterion C_t^i ($i = RMSE, LS$) is the expected difference in accuracy between TAR and VAR according to the selected metric. The calculation is based on cumulative 12-month ahead prediction for industrial production growth. Expectations are calculated regressing the difference in root mean square errors and log-scores among models on a constant and their own lag (see Section 4 for details). Positive (negative) values imply that the TAR (VAR) can be expected to be relatively more accurate based on the selected accuracy metric.
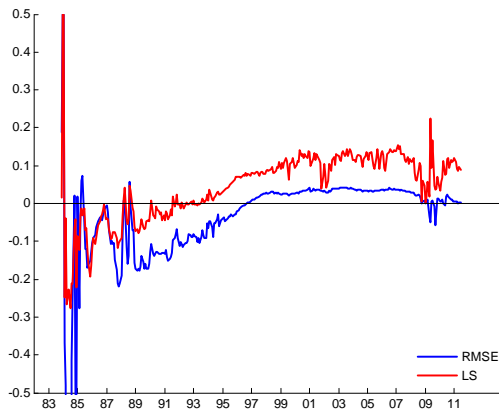

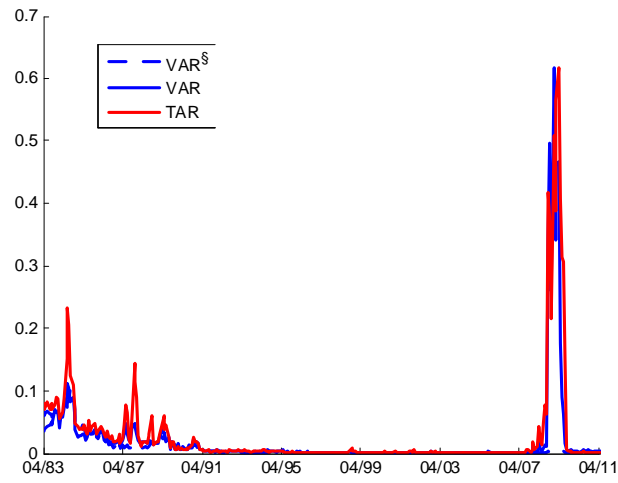Figure 11: Real-time model selection rule, consumer price inflation. See note to Figure 10.

44

Figure 12: Model-implied recession probabilities. For each model, the chart shows the probability of observing $\Sigma_{h=1}^{12} y_{t+h} < -10\%$ at any given time t, where y_t is monthly growth in industrial production.
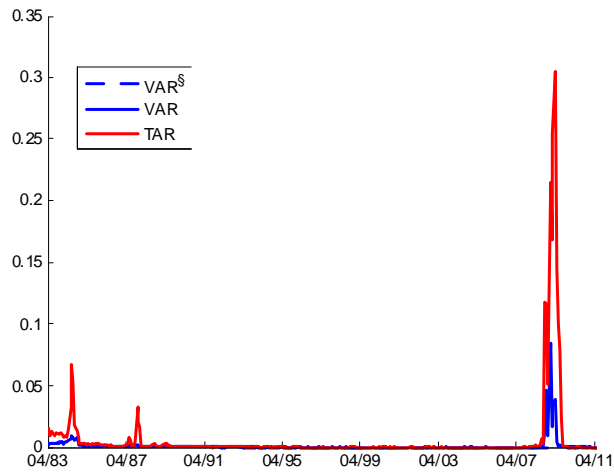

Figure 13: Model-implied recession probabilities. Probability of observing $\Sigma_{h=1}^{12} y_{t+h} < -20\%$, see notes to Figure 12.