Technical Appendix to: Financial conditions and density forecasts for US output and inflation.

Piergiorgio Alessandri Haroon Mumtaz

September 2016

Contents

A Predictive densities and financial frictions

Our work is motivated by the consideration 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 $\lim\limits_{a_t \to -\theta_t y} \mathcal{P}(a_t + \theta_t y) = -\infty$.[1] The agent's utility maximization problem thus becomes:

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

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

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

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

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

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

[1] The advantage of resorting to a penalty function to deal with the occasionally binding constraint is that the resulting model can be solved by perturbation, and, as a consequence, 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.

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]

	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

Table A.1: Decision rules with an occasionally binding borrowing constraint.

Table A.1 reports the policy functions for c_t and a_t derived from a second order perturbation of the model. All variables are expressed in deviations from steady-state. In a second order solution, the steady-state levels of consumption (resp. assets) are lower (resp. higher) relative to those obtained with a first order solution; this adjustment is represented by the "risk correction" rc in the second column of the table. The first-order terms in columns 3 to 6 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 ($a_{t-1} < 0$). Furthermore, the $\theta_{t-1}\varepsilon_t$ 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 A.1 imply for instance

[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 appropriate 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 $E(y) = 1$.

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. As long as a_t and θ_t can be observed, the central forecast from a nonlinear model consistent with the second-order approximation would 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 nonlinear model that the state-contingent implications of the shocks can 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 and 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 may 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}.

B Forecast evaluation criteria

Given a pair of models (a, b) and a variable x_t, the log-Bayes factors discussed are calculated at every point in time t as:

$$BF_t^{a,b} = \sum_{\tau=1}^{t} \left[\log(LS_\tau^a) - \log(LS_\tau^b)\right],$$

where $LS_\tau^i = p\left(x_\tau^o \,|Y_{\tau-1}, i\right)$ is the predictive density generated by model i for x_t, evaluated at the actual observation x_t^o. The tests reported and the forward-looking model selection criteria discussed in Section 6 of the paper are based on the work of Giacomini and White (2006).[4] Of interest in this case is the null hypothesis that the difference in accuracy across models at time $t+\tau$ cannot be predicted using a generic time-t information set. Given two competing forecasts for $Y_{t+\tau}$, f_t and g_t, and a loss function $L_{t+\tau}$, this can be written as:

$$H_0 : E\left[L_{t+\tau}\left(Y_{t+\tau}, f_t\right) - L_{t+\tau}\left(Y_{t+\tau}, g_t\right)|I_t\right] \equiv E\left[\Delta L_{m,t+\tau}\,|I_t\right] = 0,$$

where I_t is an information set available in t. We calculate the test for both $L = RMSE$ and $L = LS$. The unconditional tests are based on $I_t = \{\emptyset, \Omega\}$, and are calculated using only the sample mean of $\Delta L_{m,t+\tau}$. Conditional tests and decision criteria can be constructed in a number of ways. As in Giacomini and White (2006), we consider the case where I_t coincides with the information set used to generate the forecasts ($I_t = [Y_1...Y_t]$), and use a test statistic based on the sample mean of $h_t\Delta L_{m,t+\tau}$, where h_t=[1 $\Delta L_{m,t}$].[5] Intuitively, the statistic deviates from zero when the average discrepancy $\Delta L_{m,t}$ is non zero and/or it can be predicted using its own lags. The distribution of both conditional and unconditional tests are derived for the case of fixed (e.g rolling) estimation windows. Since we use an expanding window, our pvalues are only indicative.

Any persistence in $\Delta L_{m,t}$ can also be exploited to set up a decision rule that identifies at every t the model/forecast that is expected to perform better in $t+\tau$. For every time window $[1, ..., T_0 + k - \tau]$ (where T_0 = 1983.04, T = 2012.08 and $k = \tau, .., T-\tau$), we estimate a regression of the form $\Delta L_{m,t+\tau} = \delta' h_t + \varepsilon_t$, where ε_t is a white noise disturbance, and use the sample-specific OLS coefficients $\hat{\delta}_k$ to calculate the decision criterion $C_t \equiv \hat{\delta}_k' h_t$. We again calculate the criterion for both

[4] Tests and decision criteria are documented further in Section E and F of this annex.

[5] Giacomini and White use the same *test function* h_t to analyse numerically size and power properties of the test. The matlab code that implements the test is available on Raffaella Giacomini's webpage, http://www.homepages.ucl.ac.uk/~uctprgi/ .

loss functions, $L = RMSE$ and $L = LS$. The rule is simple: use f_t (respectively g_t) if and only if $C_t < 0$ (respectively $C_t > 0$). This rule picks at every t the model that is expected to generate a lower loss in $t + \tau$ (one could of course use a non-zero threshold $c > 0$ and pick model f if and only if the expected gain exceeds the threshold, $\delta' h_t > c$). We refer the reader to Section 4 of Giacomini and White (2006) for further details.

As mentioned in Section 3 of the paper, all log-scores are estimated using kernel methods. The reason for this choice is that, although the densities generated by the linear VARs are all Gaussian by construction, the densities generated by the TAR model (and some of the models introduced in the discussion in Section 7) can have strong non-normal features when the forecasting horizon h is greater than one month. For $h > 1$ the regimes may change endogenously over the horizon, so the time-$t + h$ density is generally a mixture of Gaussians. The duration of a given regime is state-dependent, and more than one switch may occur over the forecasting horizon, so characterizing these distributions analytically is difficult. Given that normal approximations cannot be used in these cases, for consistency we rely on kernel methods throughout the analysis. We use a normal kernel and estimate the densities over evenly-spaced grids of 100^n points, where n is the number of variables included in the distribution. Sensitivity analysis shows that (i) for the linear VARs this method essentially delivers the same log-scores of a parametric Gaussian approximation; and (ii) using finer grids yields virtually identical results.

C Additional impulse-responses from the TAR

Below we report additional impulse-response functions for the TAR model introduced in Section 4.2 of the paper. All responses are calculated by simulation methods following Koop et al. (1996). Financial shocks are identified recursively as explained in the Section 5 of the paper. Each of the figures below has three columns that display respectively the responses conditional on normal times (column one), the responses conditional on a crisis (column two), and the differences between the two (column three). We present the results from four experiments. In the first two experiments, we estimate the model on the entire 1973-2012 sample and simulate an increase in f_t of either one standard deviation (Figure C.1) or 0.1 units (Figure C.2). These figures are directly comparable to those presented in Section 5 of the paper and provide further information on the relative contribution of volatilities and transmission mechanisms to the asymmetry between regimes discussed in the text. In the last two experiments the analysis is replicated using a shorter estimation sample that ends in December 2007, excluding the Great Recession. The financial shock is again normalized either to one standard deviation (Figure C.3) or to 0.1 units (Figure C.4). The comparison between Figure C.1 and Figure C.3 reveals that, although the Great Recession significantly reinforced the evidence of nonlinearity in the data, the asymmetry is also clearly present in the pre-2008 sample. Figures C.2 and C.4 show that shifts in volatilities and transmission mechanisms both contribute to the larger responses observed in the crisis regimes. The relative importance of these two factors is roughly the same in the two samples, and the bulk of the asymmetry stems from an increase in the volatility of the financial shocks in periods of distress.

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. However, the mean of the posterior for the threshold z^* is around 0.25, so a 0.1 unit shock is sufficiently small to make the transition from regime one to regime two extremely unlikely. Hence, the plots in the left column of these figures can be taken to represent 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 f_t (which corresponds to five standard deviations in normal times) for instance, output falls by 5% in regime two and 3.8% in regime one.

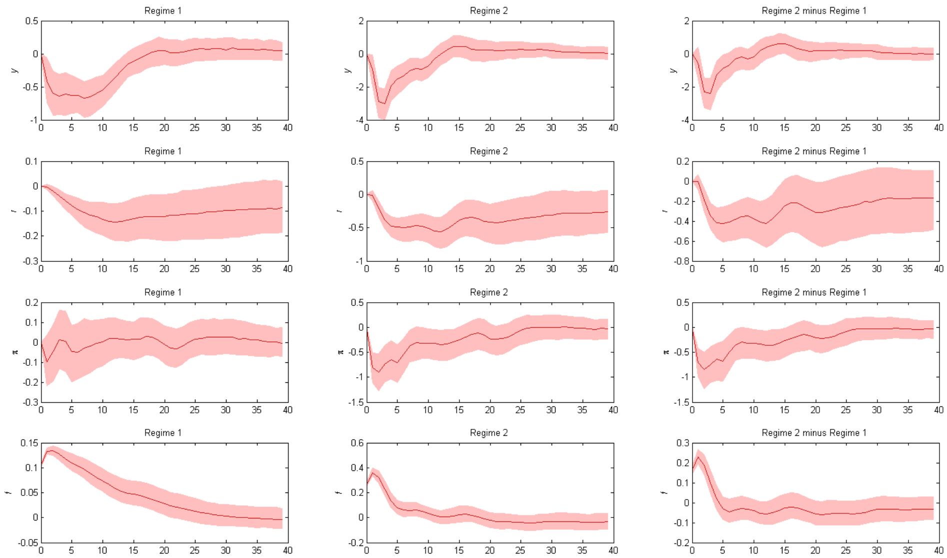

Figure C.1: Full sample TAR estimates, one standard deviation financial shock.

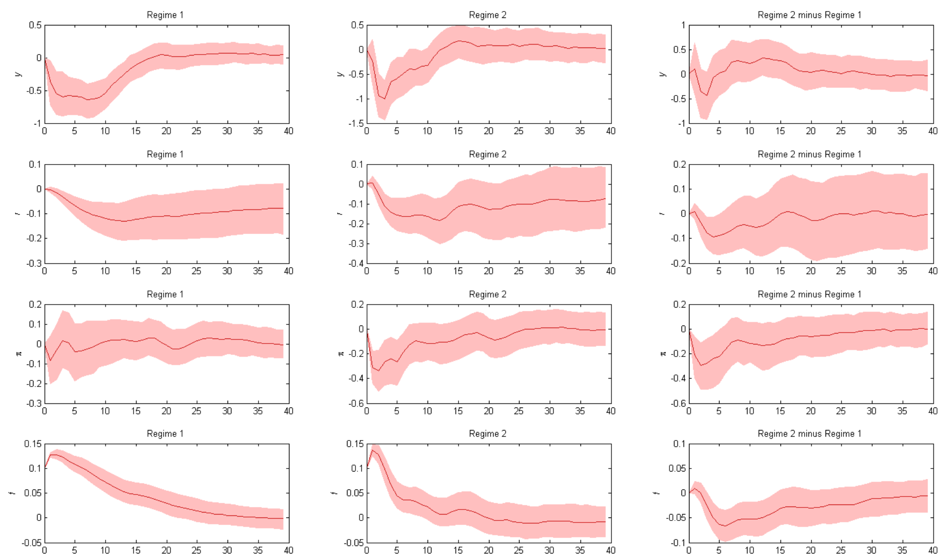

Figure C.2: Full sample TAR estimates, normalized 0.1 unit financial shock.

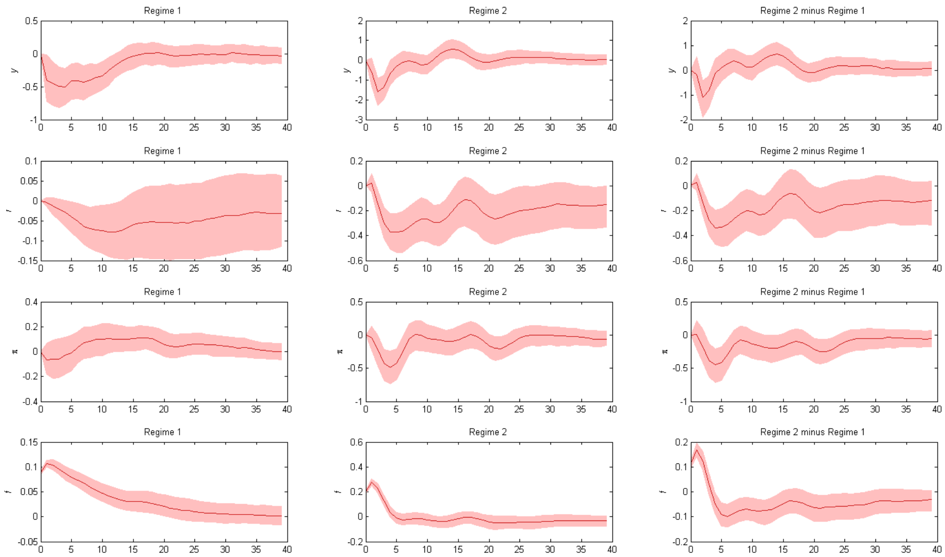

Figure C.3: Pre-crisis TAR estimates, one standard deviation financial shock.

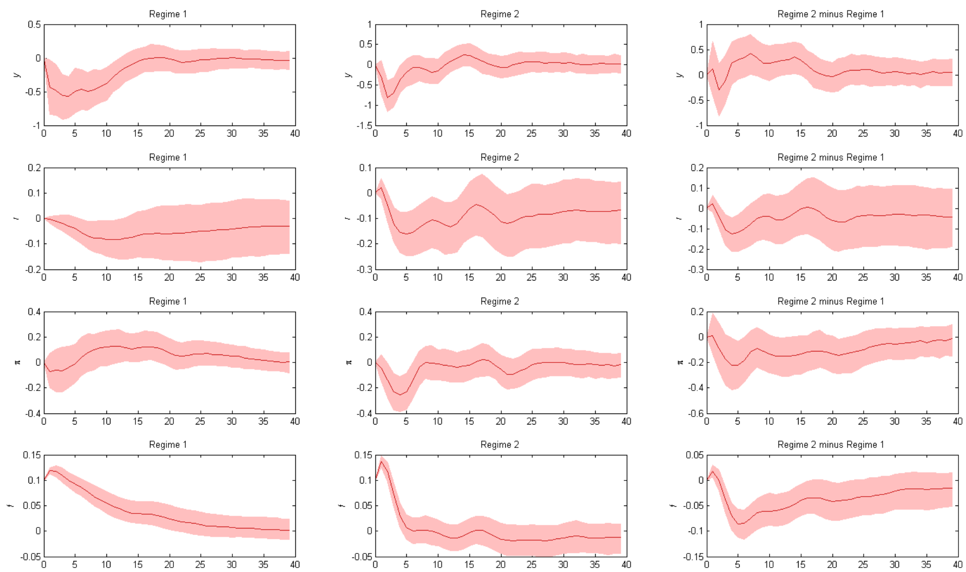

Figure C.4: Pre-crisis TAR estimates, normalized 0.1 unit financial shock.

D Model accuracy in financial distress periods

Figure E.1 shows the probability of being in regime 2 (i.e. the financial crisis) based on a recursive estimation of the TAR_f model described in Section 4.3 of the paper. Table E.1 reports the mean root mean square errors (RMSE) and log-scores (LS) generated by the models examined in the paper over these subsamples (the corresponding full-sample estimates are shown in Table 1 in the paper). The crises that fall within the 1984-2012 forecast evaluation sample are June–August 1984, November 1987–January 1988, September 1988, January–March 1989, January 2008–August 2009.

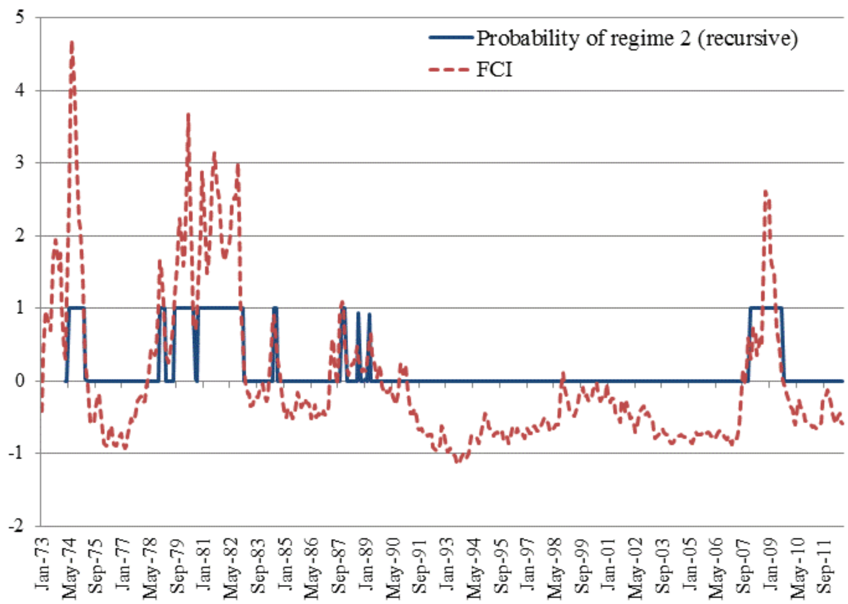

Figure D.1: Recursive estimate of the financial crisis regime from the TAR_f model.

		RMSE				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
VAR_f	y	0.638	0.773	0.945	1.033	-4.697	-3.729	-4.217	-4.981
	r	0.309	0.555	0.819	1.272	-0.709	-1.470	-1.816	-2.204
	π	0.297	0.418	0.462	0.476	-4.515	-5.641	-4.502	-2.919
	f	0.297	0.537	0.770	0.966	-2.276	-2.429	-2.854	-2.018
VAR	y	1.322	1.327	1.171	1.045	-1.198	-1.700	-4.072	-11.078
	r	0.853	0.980	1.056	1.020	0.014	-0.013	-0.067	-0.061
	π	1.019	0.968	0.972	0.935	0.043	-1.073	0.817	0.348
	f								
VAR_σ	y	1.302	1.277	1.129	1.015	-1.263	-1.374	-5.207	-17.550
	r	0.857	0.940	0.966	0.936	0.139	-0.048	0.003	-0.072
	π	0.999	0.953	0.974	0.934	1.634	2.616	1.660	0.440
	f								
VAR_f^σ	y	0.990	1.042	1.014	0.983	-4.359	-0.189	-0.961	0.507
	r	0.808	0.913	0.934	0.925	-0.887	-0.072	0.127	-0.002
	π	0.961	0.933	0.952	0.948	1.640	2.628	1.571	0.199
	f	0.884	0.869	0.850	0.869	1.878	1.190	1.277	0.443
TAR_f	y	1.136	1.100	1.067	1.059	0.702	-0.015	0.273	0.718
	r	1.450	1.629	1.634	1.427	-0.378	-0.265	-0.240	-0.151
	π	1.114	1.115	1.106	1.092	1.190	1.285	0.867	-0.159
	f	1.150	1.239	1.236	1.189	1.725	0.953	0.996	0.245
TAR_f^σ	y	1.025	1.039	1.011	0.996	0.329	0.359	0.201	0.914
	r	0.847	0.917	0.951	0.935	0.109	0.105	0.101	0.080
	π	0.972	0.979	1.002	1.015	0.934	-1.911	-0.641	-0.166
	f	0.903	0.892	0.868	0.867	-4.083	-2.176	-3.743	0.251

Table D.1: Average forecasting accuracy of the model during crisis episodes.

E Tests of equal predictive performance

Tables E.1 and Tables E.2 report the outcome of a set of pair-wise tests of equal predictive ability for the models that appear in Table 1 of the paper. For each model ($VAR, VAR^{\sigma}, VAR_f^{\sigma}, TAR_f, TAR_f^{\sigma}$) and each evaluation criterion (RMSE or log-scores, i.e. LS), the tables report the p-values for the null hypothesis that the model under examination is as accurate as the benchmark VAR^f model. Table E.1 focuses on unconditional tests while Table E.2 is based on conditional tests that check whether the discrepancy across models can be predicted based on its recent history.(see annex B for details). The p-values are purely indicative because the estimates are obtained recursively rather than with a rolling scheme.

		RMSE				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
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								
VAR^{σ}	y	0.615	0.667	0.703	0.796	0.968	0.133	0.082	0.074
	r	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.306
	π	0.057	0.000	0.000	0.000	0.037	0.918	0.019	0.000
	f	–	–	–	–	–	–	–	–
VAR_f^{σ}	y	0.004	0.188	0.871	0.475	0.443	0.000	0.268	0.037
	r	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.052
	π	0.014	0.000	0.000	0.000	0.037	0.082	0.022	0.000
	f	0.281	0.000	0.000	0.015	0.000	0.033	0.001	0.003
TAR_f	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
TAR_f^{σ}	y	0.002	0.020	0.176	0.308	0.552	0.059	0.010	0.000
	r	0.000	0.000	0.000	0.005	0.031	0.000	0.000	0.013
	π	0.132	0.025	0.023	0.002	0.065	0.873	0.095	0.000
	f	0.877	0.011	0.004	0.008	0.001	0.000	0.000	0.000

Table E.1: P-values, unconditional tests.

		RMSE				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
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	–	–	–	–	–	–	–	–
VAR_σ	y	0.615	0.667	0.703	0.796	0.968	0.133	0.082	0.074
	r	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.306
	π	0.057	0.000	0.000	0.000	0.037	0.918	0.019	0.000
	f	–	–	–	–	–	–	–	–
VAR_f^σ	y	0.004	0.188	0.871	0.475	0.443	0.000	0.268	0.037
	r	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.520
	π	0.014	0.000	0.000	0.000	0.037	0.082	0.022	0.000
	f	0.281	0.000	0.000	0.015	0.000	0.033	0.001	0.003
TAR_f	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
TAR_f^σ	y	0.002	0.020	0.176	0.308	0.552	0.059	0.010	0.000
	r	0.000	0.000	0.000	0.005	0.031	0.000	0.000	0.013
	π	0.132	0.025	0.023	0.002	0.065	0.873	0.095	0.000
	f	0.877	0.011	0.004	0.008	0.001	0.000	0.000	0.000

Table E.2: P-values, conditional tests.

F Bayes factors and model selection criteria

Figure F.1 shows the log-Bayes factors associated to the 12-month ahead predictions generated by VAR^{σ}, VAR_f^{σ}, TAR_f and TAR_f^{σ} for industrial production growth (left panel) and CPI inflation (right panel). Figure F.2 and F.2 show the forward-looking model selection criteria based on Giacomini and White (2006). We focus in this case on a direct comparison between the performances of TAR_f and VAR_f^{σ}. The criteria are calculated regressing the differences among the models' RMSE or LS on their own history, with a view to anticipating which model might be expected to perform better in the near future (see part B of this annex). Figure F.2 and F.3 illustrate the results based respectively on 6-month and 12-month ahead predictions. The normalization is such that, for all variables and accuracy metrics, positive values imply that TAR_f is expected to outperform VAR_f^{σ} over the following 6 or 12 months.

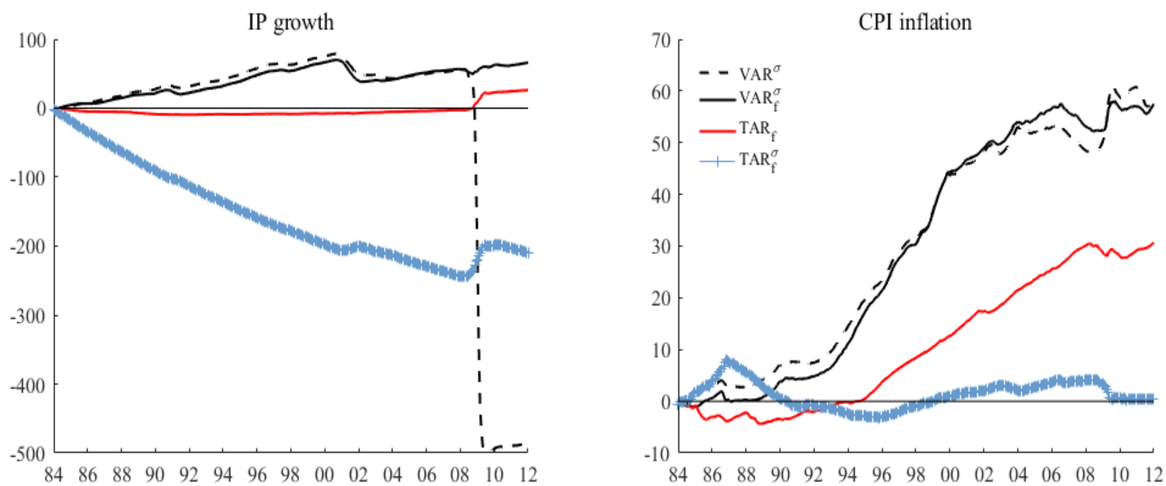

Figure F.1: Log-Bayes factors based on 12-month ahead forecasts for industrial production growth (left panel) and CPI inflation (right panel).

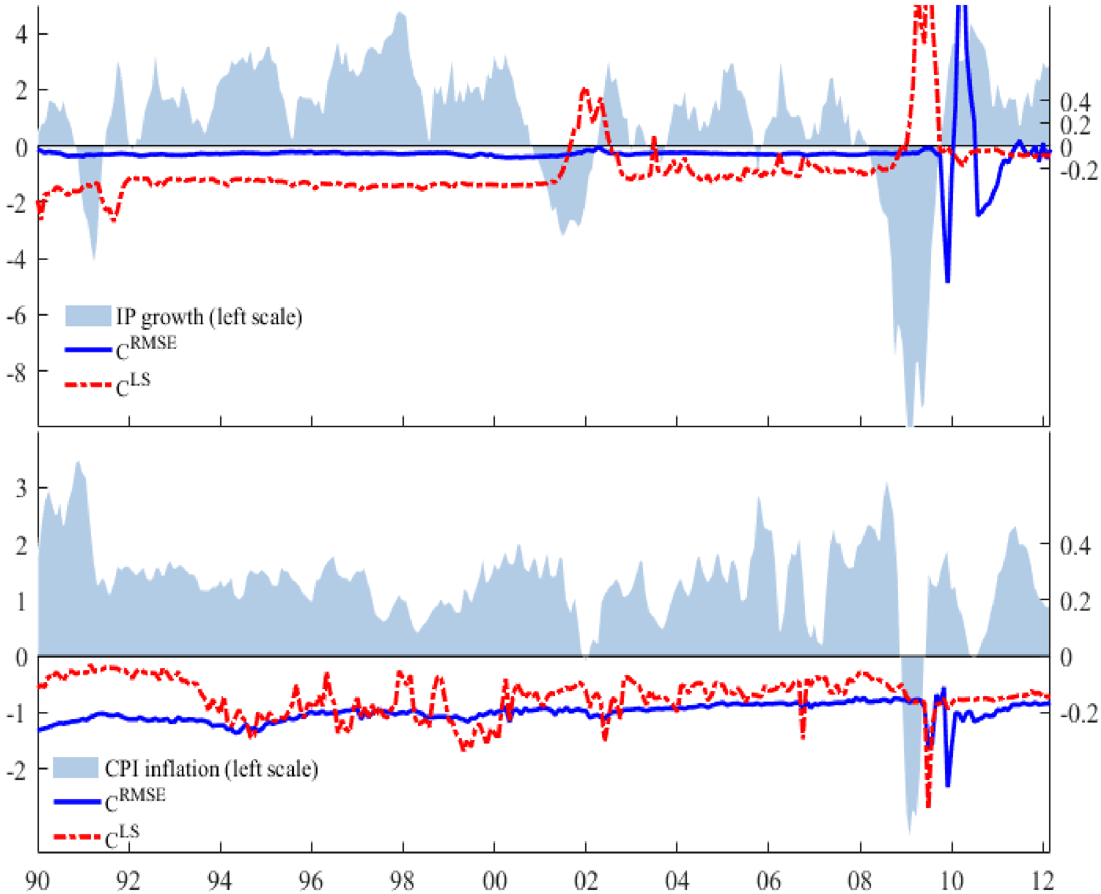

Figure F.2: Giacomini-White model selection criteria based on 6-month ahead forecasts for industrial production growth (top panel) and CPI inflation (bottom panel). The criteria are calculated using either the RMSEs (blue line) or the LS (red line) differentials between TAR_f and VAR_f^{σ}. Positive values indicate higher expected accuracy for TAR_f.

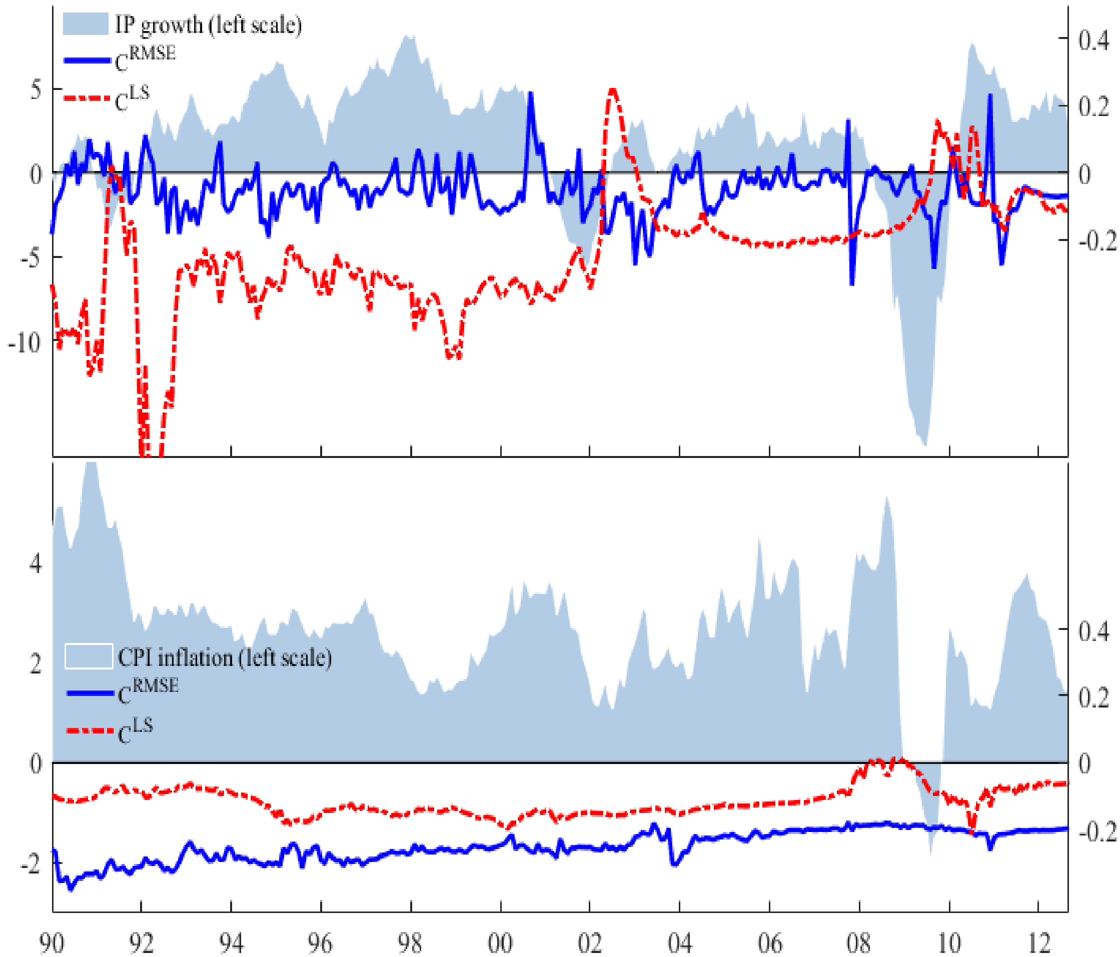

Figure F.3: Giacomini-White model selection criteria based on 12-month ahead forecasts for industrial production growth (top panel) and CPI inflation (bottom panel). The criteria are calculated using either the RMSEs (blue line) or the LS (red line) differentials between TAR_f and VAR_f^{σ}. Positive values indicate higher expected accuracy for TAR_f.

G Comparisons based on the Excess Bond Premium

Table G.1 reports root mean square errors (RMSE) and log-scores (LS) for an alternative specification of the models where the Financial Condition Index is replaced by the Excess Bond Premium of Gilchrist and Zakrajsek (2012). Data and forecasting method are as described in Section 3 of the paper except that in this case the sample ends in December 2010. We report the results for the VAR_f and TAR_f models analyzed in the paper plus the alternative Markov-switching model discussed in Annex I ($MSVAR$).

		$RMSE$				LS			
		1M	3M	6M	12M	1M	3M	6M	12M
VAR_f	y	0.470	0.544	0.574	0.599	-3.665	-3.219	-3.109	-3.044
	r	0.238	0.482	0.790	1.223	-0.704	-1.423	-1.824	-2.220
	π	0.184	0.234	0.258	0.284	-2.594	-2.621	-2.329	-2.218
	ebp	0.175	0.234	0.285	0.355	-0.898	-0.987	-1.427	-1.751
TAR_f	y	0.467	0.528	0.559	0.586	-3.576	-3.218	-3.053	-2.940
	r	0.229	0.428	0.695	1.114	-1.149	-1.286	-1.772	-2.167
	π	0.186	0.236	0.258	0.283	-2.480	-2.484	-2.239	-2.169
	ebp	0.174	0.233	0.278	0.339	-0.198	-0.610	-0.463	-0.902
$MSVAR$	y	0.462	0.531	0.567	0.594	-3.569	-3.080	-3.350	-4.637
	r	0.165	0.344	0.604	1.133	0.001	-0.994	-1.626	-2.517
	π	0.179	0.223	0.245	0.275	-2.535	-2.455	-2.606	-2.202
	ebp	–	–	–	–	–	–	–	–

Table G.1: Forecast evaluation for Excess Bond Premium specifications.

H Additional models and opinion pools

This annex collects details on the forecasting performance of the additional models and model combinations considered in the discussion of Section 7. VAR^{Large} (table H.1) is a large linear VAR which includes the four variables used in the main analysis plus the following (for each variable we provide the FRED database identifier in brackets if available): civilian employment (CE16OV) and unemployment rate (UNRATE), average weekly hours worked in the manufacturing sector (AWHMAN), M2 money stock (M2SL), housing starts (HOUST), the Reuters/Jefferies CRB spot commodity price index, and the NAPM-ISM Purchasing Managers Index. TAR^{Vol} (table H.2) has the same form as the TAR model employed in the paper, and is estimated on the same data, except that in this case only the residual covariance matrix is allowed to change across regimes.

In tables H.1 and H.2 we use a dagger (†) or a double dagger (‡) to highlight cases (i.e. variables, criteria and horizons) for which the model under consideration outperforms respectively VAR_f or TAR_f, i.e. the linear and threshold VARs analyzed in the paper. Table H.3 illustrates the forecasting performance of the opinion pools examined in the discussion of Section 7. We evaluate two types of pools: $Pool_1$ includes the linear VARs only, while $Pool_2$ also includes the TAR. In each case we consider a naïve weighting scheme that attaches constant equal weights to all models and, alternatively, the optimal weighting scheme of Geweke and Amisano (2011) and Geweke and Amisano (2012), updating the weights recursively at each point in time. The table reports the average log-scores based on 12-month ahead predictive densities, and stars are used to identify the best model or model combination for each variable.

	RMSE				LS			
	1M	3M	6M	12M	1M	3M	6M	12M
y	0.443†‡	0.500†‡	0.536†‡	0.571†‡	-3.517†	-3.136†‡	-3.005†	-2.997
r	0.204	0.382	0.592†	0.950†	-0.596†	-1.307†	-1.680†	-2.065†
π	0.174 ‡	0.223	0.244	0.268	-2.586	-2.746	-2.351	-2.218
f	0.115	0.231	0.342	0.453	-0.069	-0.877	-1.017	-1.147

Table H.1: Forecast evaluation for VAR^{Large}

	RMSE				LS			
	1M	3M	6M	12M	1M	3M	6M	12M
y	0.454 ‡	0.520	0.551	0.577 ‡	-4.274	-3.994	-3.794	-3.582
r	0.152†‡	0.313†‡	0.524†‡	0.884†‡	-0.062†	-0.849†	-1.369†	-1.945†‡
π	0.169†‡	0.214†‡	0.233†‡	0.255†‡	-2.558†	-2.771	-2.420	-2.170†
f	0.096†‡	0.174†‡	0.247†‡	0.334†‡	-0.162	-0.413†	-0.790†	-1.183

Table H.2: Forecast evaluation for TAR^{Vol}

	VAR	VAR_f	TAR_f	$Pool_1$ (VAR, VAR_f)		$Pool_2$ (VAR, VAR_f, TAR_f)	
				Equal weights	Optimal weights	Equal weights	Optimal weights
y	-3.948	-2.964	-2.885*	-3.139	-2.962	-3.197	-2.938
r	-2.118	-2.101	-1.999*	-2.002	-2.110	-2.075	-2.052
π	-2.137	-2.171	-2.080	-2.147	-2.130	-2.325	-2.069*
f	–	-1.130	-0.717	–	–	–	–

Table H.3: Forecast evaluation, baseline models versus opinion pools. Average log-scores based on 12-month ahead forecasts.

I Markov switching VAR

As an alternative to the TAR_f model described in Section 4.2 of the paper, we consider a more flexible Markov-Switching VAR with endogenous transition probabilities. In this case the strength of the link between financial conditions and changes in economic regimes can be estimated from the data, and the model allows for (and can help detecting) limiting cases where the transition probabilities are unrelated to financial conditions. The Markov switching VAR has the following form:

$$Y_t = c_{S_t} + \sum_{j=1}^{P} B_{j,S_t} Y_{t-j} + \Omega_{S,t}^{1/2} e_t, \quad e_t \tilde{} N(0,1), \tag{6}$$

where $Y_t = (y_t \; r_t \; \pi_t)$ collects output, inflation and federal funds rate and $S_t = \{0, 1\}$ denotes the unobserved regime, which is assumed to follow a first order Markov Chain. The transition probabilities are time-varying and depend on the financial indicator f_t, which in the baseline case is again taken to be the Financial Condition Index:

$$\begin{pmatrix} P(f_{t-1}) & 1 - Q(f_{t-1}) \\ 1 - P(f_{t-1}) & Q(f_{t-1}) \end{pmatrix} \tag{7}$$

The process for the transition probabilities can be represented as a Probit model (see Filardo and Gordon (1998)) defined via the following latent variable y_t^*.

$$\begin{aligned} &S_t = 1 \iff y_t^* \geq 0 \\ &y_t^* = \lambda_0 + \gamma_1 f_{t-1} + \lambda_1 S_{t-1} + v_t, v_{it} \tilde{} N(0,1) \end{aligned} \tag{8}$$

Therefore $P(f_{t-1}) = \Pr(S_t = 0 \backslash S_{t-1} = 0)$ is given by:

$$\begin{aligned} \Pr(S_t = 0 \backslash S_{t-1} = 0) &= \Pr(v_t < -\lambda_0 - \gamma_1 f_{t-1} - \lambda_1 S_{t-1}) \\ &= \Phi(-\lambda_0 - \gamma_1 f_{t-1}), \end{aligned} \tag{9}$$

and $Q(f_{t-1}) = \Pr(S_t = 1 \backslash S_{t-1} = 1)$ is given by:

$$\begin{aligned} \Pr(S_t = 1 \backslash S_{t-1} = 1) &= \Pr(v_t \geq -\lambda_0 - \gamma_1 f_{t-1} - \lambda_1 S_{t-1}) \\ &= 1 - \Phi(-\lambda_0 - \gamma_1 f_{t-1} - \lambda_1 S_{t-1}), \end{aligned} \tag{10}$$

where $\Phi(.)$ is the standard normal CDF. Note that λ_0 and λ_1 are regime specific constants. If $\gamma_1 = 0$, then we are back to the standard fixed transition probability Markov-switching specification. More generally, the coefficient γ_1 determines the impact of the financial indicator f_t on the transition probabilities. Since the normal distribution is symmetric, $\Pr(S_t = 0 \backslash S_{t-1} = 0) = \Phi(-\lambda_0 - \gamma_1 f_{t-1}) = 1 - \Phi(\lambda_0 + \gamma_1 f_{t-1})$. This implies that if $\gamma_1 < 0$ then an increase in f_{t-1} increases $\Pr(S_t = 0 \backslash S_{t-1} = 0)$. We assume that the prior on the VAR parameters in the two regimes is of the natural conjugate form defined in Section 4 of the paper. The prior on the parameters of the transition probability equation is assumed to be normal and defined as:

$$p\begin{pmatrix} \lambda_0 \\ \gamma_1 \\ \lambda_1 \end{pmatrix} \tilde{} N\left(\begin{pmatrix} -2 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}\right) \tag{11}$$

The prior mean implies that, absent the influence of f_{t-1}, the regimes are persistent. A detailed description of the Gibbs sampler for this model can be found in Amisano and Fagan (2013). Here we sketch the basic steps. Given a draw for the regime S_t, equation (6) collapses to a sequence of BVAR models with conditional posterior given by equations (4) and (5) of the paper. Given a draw for the VAR parameters and the transition probabilities, the unobserved regime S_t is drawn using the multi-move algorithm described in Kim and Nelson (1999). Given a draw for $S_t, \lambda_0, \lambda_1$ and γ_1, the latent variable y_t^* can be drawn from a truncated normal distribution. That is $y_t^* \tilde{} N_{I<0}(\mu, \tau)$ if $S_t = 0$ and $y_t^* \tilde{} N_{I>0}(\mu, \tau)$ if $S_t = 1$, where $I < 0$ denotes truncation below zero and $I > 0$ denotes truncation above zero. Note that $\mu = \lambda_0 + \gamma_1 f_{t-1} + \lambda_1 S_{t-1}$ while $\tau = 1$ for identification. The transition probabilities can be easily calculated using equations (9) and (10). Given a draw for y_t^*, equation (8) is a linear regression with an intercept λ_0, an intercept interacted with S_{t-1} with coefficient λ_1 and a regressor f_{t-1}. The conditional posterior for these parameters is of standard form and given by a normal density. We employ 20,000 iterations of the Gibbs sampler discarding the first 15,000 as burn-in. In order to deal with the label switching problem, we impose the condition that the implied unconditional mean of y is lower in regime 0. The forecast density for this model is defined as$G(Y_{t+K} \setminus Y_t) = \int G(Y_{t+K} \setminus Y_t, \Gamma) \times G(\Gamma \setminus Y_t)\, d\Gamma$, where $\Gamma = \{B_{S_{t+K}}, \Omega_{S_{t+K}}, P_{t+K}(f_{t+K-1}), Q_{t+K}(f_{t+K-1}), S_{t+K}\}$. We use the BVAR described in section 4.1 of the paper to obtain the forecast density for f_t, which is assumed to be an exogenous indicator in the MSVAR model. Given forecast values for f_t and the m_{th} Gibbs draw $\lambda_{0,m}, \lambda_{1,m}$ and $\gamma_{1,m}$, the transition probabilities can be projected forward using equation (8). Given $P_{t+K,m}, Q_{t+K,m}$ the prediction step of the Hamilton (1989) filter is used to estimate $\Pr(S_{t+K} = i)_m$ for $i = 0, 1$ and the m_{th} draw from the forecast density is calculated as a weighted average of equation

(6) iterated forward using $B_{0,m}, \Omega_{0,m}$ and $B_{1,m}, \Omega_{1,m}$.

The support of the estimated posterior for γ_1 is entirely negative and centered around -1.2, suggesting that financial conditions are indeed an important driver behind changes in economic regimes. The regimes appear to be less persistent than those estimated by the TAR. Probability integral transforms and coverage ratios for this model are qualitatively similar for the two models (see Annex **??**). Table I.1 below reports average root mean square error and log-scores for all variables and horizons. A comparison between these figures and those in Table 1 of the paper shows that the model is marginally better than the TAR in terms or root mean square errors, but worse in terms of log-scores. To shed light on this result, in figure I.1 we plot log-Bayes factors (i.e. cumulative log-score differences, see Annex B) that compare the performance of the $MSVAR$ to that of the VAR and TAR models over time. In the case of output (y_t, left panel), the cause of the bad performance of the $MSVAR$ is clear. This model actually dominates VAR and TAR until about 2008, but the gain is entirely wiped out at the end of the sample because the model fails to predict the Great Recession. Recall that in the $MSVAR$ the financial indicator can cause changes in regimes (through equation (8)) but it does not directly affect the dynamics of the economy (because it does not enter equation (6)). Hence, this result essentially confirms that without accounting for a direct macro-financial linkage anticipating the Great Recession would have been impossible irrespective of whether one was relying on a plain linear model or a more flexible Markov-switching specification. In the case of interest rates (r_t, middle panel) and inflation (π_t, right panel), figure I.1 shows that the performance of the $MSVAR$ is broadly comparable to that of the VAR but significantly worse than that of the TAR. The comparison with VAR delivers a negative final verdict as of 2012, but this is largely due to the bad performance of $MSVAR$ during the Great Recession: the factors actually display large swings over the sample period, which cautions against concluding too confidently that $MSVAR$ should be discarded in favour of the linear model. The comparison with TAR is very different. Although the Great Recession appears to take its toll on the $MSVAR$ in this case as well, the steady decline in the factors long before 2008 suggests that this model is indeed generally less accurate, possibly because of its heavier parameterization.

	RMSE				LS			
	1M	3M	6M	12M	1M	3M	6M	12M
y	0.456	0.527	0.564	0.586	-3.709	-3.051	-3.382	-4.674
r	0.154	0.312	0.534	0.970	0.050	-0.855	-1.456	-2.190
π	0.174	0.218	0.237	0.261	-2.534	-2.516	-2.694	-2.170
f	-	-	-	-	-	-	-	-

Table I.1: Root mean square errors and log scores for the $MSVAR$ model.

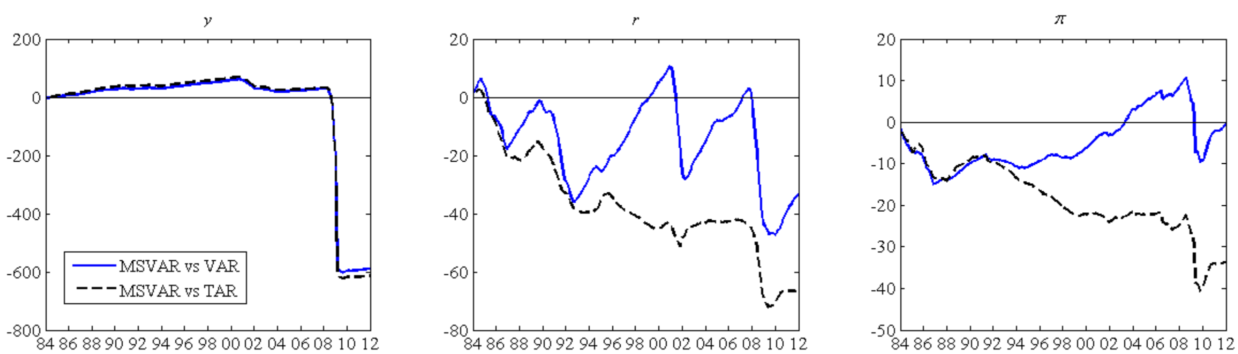

Figure I.1: Log-Bayes factors for the MSVAR model.

References

Amisano, G. and G. Fagan (2013). Money growth and inflation: A regime switching approach. *Journal of International Money and Finance 33*, 118–145.

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

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). Business cycle durations. *Journal of Econometrics 85*(1), 99–123.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130–141.

Geweke, J. and G. Amisano (2012). Prediction using several macroeconomic models. *European Central Bank. Frankfurt*.

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

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

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

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

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.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics 74*(1), 119–147.

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