Appendix for

FEEDBACKS: FINANCIAL MARKETS AND ECONOMIC ACTIVITY

For Online Publication

Markus K. Brunnermeier, Darius Palia, Karthik A. Sastry, and Christopher A. Sims
CONTENTS

1. Choice of Prior Distributions 2
   1.1. Shock Impacts $A_0$ and Relative Variances $\Lambda_i$ 2
   1.2. Reduced Form Coefficients $B_j$ 2
   1.3. “Fat-tail shocks” $\xi_{i,t}$ 3
2. Posterior Sampling Method 3
   2.1. First Block ($P[\theta_1 | \theta_3, Y]$) 4
   2.2. Second Block ($P[\theta_2 | \theta_1, \theta_3, Y]$) 6
   2.3. Third Block ($P[\theta_3 | \theta_1, \theta_2, Y]$) 6
3. Computational Details and Algorithm Convergence 7
   3.1. Execution and Timing 7
   3.2. Burn-in 8
   3.3. Convergence and Effective Sample Size 8
   3.4. Marginal Likelihood Estimates 11
4. Construction of Credit Data 12
   4.1. Correction for Measurement Breaks 13
   4.2. Comparison with Quarterly Data 14
5. Large Residuals in the Main Model 16
6. Robustness of the Main Model 18
   6.1. Different Error Distributions 18
   6.2. Shortened Sample Period 19
   6.3. Triangular Normalization 22
   6.4. Long-term Growth Rates 28
   6.5. Non-linear Transformation 28
7. Models with Fewer Variables 30
   7.1. A Small VAR Model 30
   7.2. Monte Carlo Exercise 30
8. Models with Quarterly Data 32
References 36
1. CHOICE OF PRIOR DISTRIBUTIONS

As described in the main text, the model with Gaussian errors has \( n^2 \) free parameters in \( A_0 \), \((M - 1)n\) free parameters in the \( \Lambda_m \), and \( n^2 p \) free parameters in the \( A_j \), where \( n \) is the number of observed variables, \( M \) the number of variance regimes, \( T \) the number of observations excluding initial conditions, and \( p \) the number of lags. The mixture of normal models (including the baseline \( t \) model) have an additional \( nT \) parameters.

1.1. Shock Impacts \( A_0 \) and Relative Variances \( \Lambda_i \). The Gaussian prior for \( A_0 \), treating each element as independent with standard deviation 200 and means of zero off the diagonal and 100 on the diagonal, implicitly scales each equation to an expected prior standard deviation of 0.01. This is a standard assumption for log macro time series, and potentially a slight over-estimate for interest rates (expressed in decimal terms, with 0.01 as 1 percent).

The Dirichlet(2) prior on \( \lambda_{i} / M \), a multivariate generalization of a Beta distribution, centers the model around equal variance in each period and enforces the normalization that the average variance equal one.

1.2. Reduced Form Coefficients \( B_j \). For priors on the reduced form coefficients, we implement dummy observations following Sims and Zha (1996).

First, a set of “Minnesota prior” dummy observations centers belief around independent random walks in each equation, with prior precision around a mean of zero increasing for further lags. In the notation of the reference, we use a tightness of 3 and decay of 0.5.

A second set of dummy observations, a “unit root prior,” expresses belief that the variables will stay at a “mean level.” We estimate the prior mean as the sample mean from the lagged observations, which are not used on the left-hand-side of estimation. One observation expresses belief that all variables stay at the level, and another \( n \) observations express the belief that each independently stays at the level. Again, in the notation of the reference, we specify this with tightness 5 and persistence 1.

We turn these priors on the reduced form coefficients of \( B_j \) into priors on the structural coefficients \( A_j \) by multiplying the dummy data, along with the actual data, by \( A_0 \).
1.3. “Fat-tail shocks” $\xi_{i,t}$. We assume the “fat-tail shocks”\(^1\) in the normal mixture versions of the model are i.i.d. across equation and time. The former restriction is natural because, \textit{a priori}, we do not distinguish the order of equations. The latter we claim is reasonable since the regime-changing heteroskedasticity controls low-frequency volatility movements.

In the normal mixture case, we assume three normal distributions (based on fitting residuals from the standard model) with prior probabilities $\alpha_1, \alpha_2, \alpha_3$. In the $t$-distributed case, we assume a prior distribution of inverse-gamma(scale = $k/2$, rate = $2/k$) to make the errors $t$ with unit scale and $k$ degrees of freedom.

To select parameters, we estimated a maximum likelihood fit based on residuals from the Gaussian error model. We picked standard deviations .68, 1.17, and 2.55 with prior probabilities .59, .39 and .02 in the first case. The second case has degrees of freedom 5.7. On the residuals, the difference in fit between the two models was only 0.22; this was a 179 log point improvement over a model of independent, unit-scale normals.

2. POSTERIOR SAMPLING METHOD

This section describes in detail the method we use to sample the posterior in the models with normal mixture errors.\(^2\) The procedure for sampling the model with Normal errors is nested within. We separate the parameters into the following three blocks:

(1) $\theta_1 = \left\{ A_0, (A_m)_{m=1}^M \right\}$. These are the contemporaneous coefficients and their scalings.

(2) $\theta_2 = \left\{ (A_i)_{i=1}^p, (\xi_{i,t})_{i,t=1}^{n,T} \right\}$. These are the structural form coefficients and the structural shocks.

(3) $\theta_3 = \left\{ (\xi_{i,t})_{i,t=1}^{n,T} \right\}$. These are the variance adjustments (or high-frequency volatility shocks) for each equation in each time period. We can think of them as uniformly 1 (point mass prior) if errors are Gaussian.

\(^1\) The main article discusses the sense in which we could call these “parameters” or “shocks” — in a Bayesian framework, they are treated symmetrically as random variables.

\(^2\) The next section (C) describes in more detail the computational implementation of the method and its numerical convergence.
and proceed with a Gibbs sampling algorithm, a special case of the Metropolis-Hastings posterior sampling algorithm that samples from the joint posterior $P[\theta_1, \theta_2, \theta_3 \mid Y]$ by iteratively taking samples from conditional distributions.

Our approach, outlined in Table 1, is a somewhat non-standard Metropolis-in-Gibbs approach because the first step samples from a distribution integrated over the parameter block $\theta_2$. This poses no additional technical challenges for proving the validity of Gibbs sampling for delivering a random sample from the joint posterior $P[\theta_1, \theta_2, \theta_3 \mid Y]$. But it does greatly enhance the efficiency of the Metropolis-Hastings step by eliminating $n^2p$ highly dependent parameters from the updating.\(^3\) Step (3) also technically uses a marginal posterior, but that is also an exact conditional (put differently the $\epsilon_{it}$ are sufficient to define that conditional distribution).

The remainder of this section describes each block in further detail.

2.1. First Block ($P[\theta_1 \mid \theta_3, Y]$). This step requires a Metropolis update, and is the most computationally intensive of the three.

Recall the structure of the model

$$A_0y_t = \sum_{j=1}^{p} A_j y_{t-j} + C + \epsilon_t$$ \hspace{1cm} (1)

with

$$\epsilon_{i,t} \sim \text{Normal}(0, \lambda_{i,t} \xi_{i,t})$$ \hspace{1cm} (2)

A useful notation for each equation $i$ of this system is

$$Y((A_0)_i,) = X(A_+)_i + E_i$$ \hspace{1cm} (3)

where we stack all the left-hand-side data into a $T \times n$ matrix $Y$, we select the $i$th row of $A_0$ (and then transpose it into a column vector), we collect all right-hand-side variables

\(^3\)As will be illustrated below, the $nT$ parameters of $\epsilon_{it}$ are linear functions of $Y$ and the $A_i$, so they do not need to be drawn separately.
into the matrix $X$, stack the $i$th rows of the $A_i$ and constant $C$ into the vector $(A_+)_i$, and stack all the $\varepsilon_{i,t}$ into $E_i$.

The first step is to augment the observed data in $X$ and $Y$ with “dummy observations” that implement the prior. These observations are described in more detail in the first Appendix section. This gives the augmented system of equations

$$\hat{Y}((A_0)_i)' = \hat{X}(A_+)_i + E_i$$

Next we need to turn $E_t$ into Gaussian errors of unit variance. One way to represent this is to construct a matrix $\Xi_i$ with diagonal elements $\sqrt{\zeta_{i,t}}$ and a matrix $L_i$ with diagonal elements $\sqrt{\lambda_{i,t}}$ and then work with

$$\Lambda_i \Xi_i \hat{Y}((A_0)_i)' = \Lambda_i \Xi_i \hat{X}(A_+)_i + \Lambda_i \Xi_i E_i$$

Now it is straightforward to find the posterior mode, the posterior covariance matrix for the coefficients of each equation (in $(A_+)_i$), and the marginal likelihood $P[Y_t \mid Y_0, A_+, \theta_1, \theta_3]$.

The goal for our sampling is to evaluate $P[Y_t \mid Y_0, A_+, \theta_1, \theta_3] P[\theta_1]^1$, an un-normalized posterior, and apply a standard Random Walk Metropolis (RWM) algorithm. In this case, starting from some draw $\theta_{1,j}$, we generate a proposal $\theta_{1}^{(k)}$ from a Gaussian distribution with mean $\theta_1'$ and specified variance $V_{\theta_1}$, customarily estimated from an optimization to find the posterior mode (in this case, an optimization of the model with Gaussian errors) times a scale factor. Scale factors of 0.05-0.10 were successful in generating reasonable acceptance rates of about 20 to 30 percent.\(^4\)

Given the proposal, we set $\theta_{1}^{(k+1)} = \theta_{1}'$ with probability

$$\min \left\{ 1, \frac{P[Y_t \mid Y_0, A_+, \theta_1', \theta_3] P[\theta_1']}{P[Y_t \mid Y_0, A_+, \theta_1^{(j)}, \theta_3] P[\theta_1^{(j)}]} \right\}$$

else we set $\theta_{1}^{(k+1)} = \theta_{1}^{(k)}$.

\(^4\)Gelman et al. (2014) suggests that RWM jump rules for a $d$-dimensional posterior are optimally efficient with covariance matrix $c^2 \Sigma$ where $c = 2.4/\sqrt{d}$ and $\Sigma$ is an estimate of the posterior covariance matrix. For $d = 150$, this formula suggests $c^2 = 0.0384$, which is close to what we use. In this context, efficiency measures the number of MCMC draws to direct draws from the posterior to achieve a target variance. A more efficient sampler needs fewer draws to give us the same “performance” with respect to this (important) moment.
2.2. Second Block ($\mathbb{P}[\theta_2 \mid \theta_1, \theta_3, Y]$). The previous section already suggested the conditional distribution of $A_+$, which is Gaussian with a block diagonal covariance structure. More explicitly, if $\hat{X} := \Lambda_i \Xi_i \tilde{X}$ and $\hat{Y} = \Lambda_i \Xi_i \tilde{Y}((A_0)_i,)'$

$$\mathbb{P} [(A_+)_i, \mid A_0, \Lambda] \sim \text{Normal} \left( (\hat{X}' \hat{X})^{-1} \hat{X}' \hat{Y}, (\hat{X}' \hat{X})^{-1} \right) \quad (7)$$

Note that if the third block is degenerate (e.g., if we want the model with Normal shocks), we can delay this block until after completing a full sample from the first. This is computationally convenient, as we may only want to compute impulse responses for some subset (often subsampling) of the full sample of $(A_0, \Lambda)$.

In the general case, we also want to form the structural errors. In fact the object that is more useful to us will be $u_{i,t} := \varepsilon_{i,t} / \lambda_{i,t}$, and that is formed as

$$U_i = (\Xi_i)^{-1} (\hat{Y} - \hat{X}(A_+)_i,) \quad (8)$$

The implementation of this is to take the residuals from the linear regression (which is evaluated at the conditional posterior mode of $A_+$), subtract the correction from the deviation of the $A_+$ draw from that posterior mode, and then divide by the $\xi_{i,t}$ element-by-element.

2.3. Third Block ($\mathbb{P}[\theta_3 \mid \theta_1, \theta_2, Y]$). In the two cases we consider, there are simple ways to sample the posterior distributions of the $\xi_{i,t}$ with closed-form distributions. The important object, as stated previously, is $u_{i,t} := \varepsilon_{i,t} / \lambda_{i,t}$, the residual before transformation by $\xi_{i,t}$. We can observe that $\mathbb{P}[\xi_{i,t} \mid \theta_1, \theta_2] = \mathbb{P}[\xi_{i,t} \mid u_{i,t}]$, and the updates can be done in parallel for each equation and each time period.

If $\xi_{i,t}$ has an inverse-gamma distribution with shape $\alpha$ and rate $\beta$, the posterior distribution conditional on $u_{i,t}$ is IG($\alpha + 1/2, \beta + u_{i,t}^2/2$).

In the multinomial normal mixture case, which is discussed in more detail in Appendix 6, the posterior distribution places posterior probability mass

$$p_i = \frac{\alpha_i \phi(u_{i,t}/\sqrt{x_i})/\sqrt{x_i}}{\sum_{i=1}^N \alpha_i \phi(u_{i,t}/\sqrt{x_i})/\sqrt{x_i}}$$

on each of the $N$ possible values $x_i$, where $\alpha_i$ are the prior weights and $\phi(\cdot)$ evaluates the Gaussian pdf.
3. COMPUTATIONAL DETAILS AND ALGORITHM CONVERGENCE

This Appendix section describes, in some detail, the implementation of the posterior sampling method described in Appendix B. A code supplement provides software to replicate all of this paper’s empirical analysis in $\mathbb{R}$.\(^5\)

3.1. Execution and Timing. Major calculations were executed on two machines, (1) a 16-core, 2.60GHz Intel Xeon workstation and (2) a 4-core, 4.00 GHz Intel i7-4790K desktop. Evaluating a function proportional to the posterior of $A_0$ and $\Lambda$ conditional on the $\xi_{i,t}$ and integrated over the coefficients $A_+$ (in the previous section’s notation, $P[\theta_1, \theta_3 \mid Y]$) took approximately 0.20 seconds for the full model (10 variables, 6 variance regimes) and the full set of data. Each posterior draw required two likelihood evaluations (for the Metropolis-Hastings step), so we could simulate about 215,000 draws per computer per 24 hours.

Obtaining a reliable sample from the model posterior(s) proved challenging. Our main results are based off of pooling draws among several dispersed MCMC chains. We believe this approach should give conservative estimates of parameter variance and allay concerns that our algorithm sampled only a small subset of the parameter space (e.g., around one thin peak but ignoring other high density areas).

For each model, we ran two stages of MCMC with an adaptive step in the middle. Before any runs, we ran an optimization algorithm on the posterior of the normal errors model and obtained posterior mode $x_0$ and estimated inverse Hessian matrix $C_0$. Our first stage of MCMC consisted of several parallel chains started from dispersed starting points $x_0 + N(0, C_0)$, where $N(0, C_0)$ was a Gaussian random vector with zero main and covariance $C_0$. Each chain used random-walk Metropolis transition density $N(0, kC_0)$, where $k$ was adjusted to obtain about 20% acceptance on the Metropolis step (set around 0.04). In the second stage, we calculated a covariance matrix $C_1$ from the pooled draws of the first stage, picked an $x_1$ from the highest (conditional on $\xi_{i,t}$) density draw from the first stage, and set starting points $x_1 + N(0, C_1)$ and RWM transition densities $N(0, kC_1)$ tuned in the same way.

\(^5\)Documentation within the code supplement describes, in some more detail, how to use all the programs and recreate specific figures and tables.
3.2. **Burn-in.** We based our estimate of appropriate “burn-in” (discarded initial draws before sufficient convergence to the posterior) on the stability of trace plots for the log posterior density (from the Metropolis step, so conditional on \((\xi_{it})_{i,t}\) and integrated over the \((A_i)_{i \geq 1}\)) and the parameters. Figure 1 provides some example trace plots from selected parameters of one representative MCMC chain. The initial draw indexed 0 is more than 300,000 draws from the initialization. We take the lack of low-frequency trend in the log posterior density (and parameter) trace plots as some evidence that the sampler has stabilized around the highest density area of the posterior.

3.3. **Convergence and Effective Sample Size.** There are two general areas of concern for an iterative simulation posterior sample:

1. Within-sequence auto-correlation reduces the “effective” amount of independent draws.
2. The simulation sequences do not represent the target (posterior) distribution. Symptoms of this can include disagreement among chains with dispersed starting points, or significant non-stationary “drift” in certain parameters over draws.

The simplest, within-chain measure of “effective sample size” is the appropriate \(n_e\) such that \(\text{Var}[\hat{\mu}] = \text{Var}[\mu]/n_e\), where \(\hat{\mu}\) is the empirical mean and \(\sigma^2\) is the (consistently estimated) variance of each scalar parameter. The first three columns of Table 2 (“Within chain”) gives quantiles of these ESS statistics for each parameter and for each chain (i.e., so the quantiles are over (number of chains) x (number of parameters) objects). The ESS for the impulse response parameters are relatively healthy, and tend to increase linearly (if slowly) with sample size (based on incremental ESS calculations which are available upon request). Less confidence about the sampling of certain parameters of \(A_0\) and \(\Lambda\), which might also reflect relatively poor identification, is less important for the main qualitative results of our paper, so we find it less concerning.

The next three columns of Table 2 gives an “across chain” measure of ESS based on the individual chain variances of each parameter relative to the pooled variances. In this sense, it jointly addresses concerns (1) and (2). Across chain ESS is considerably lower for

---

6Hear, the impulse response parameters are the “average” response of each variable to each shock at the 1, 12, and 48 month horizons.
Figure 1. Trace plots for log posterior density (top left); the 6th diagonal entry of $A_0$, corresponding to the monetary policy equation (top right); the variance of the 6th (monetary policy) shock in the first period (bottom left); and the first auto-regressive coefficient for the 6th shock ($A_1[6,6]$, where the square brackets index the matrix). The black line plots every 1,000th draw, and the blue line is a two-sided moving average of 1,000 draws subsampled from every 10,000.

all parameters, which motivates our using pooled MCMC draws for major calculations. But ESS for the impulse responses seems relatively reasonable.

A different way of presenting the across-chain ESS measures is given in Table 3, which reports the Gelman and Rubin (1992) “potential scale-reduction factor” defined as
TABLE 2. Distribution of effective sample size (ESS) measures of posterior simulation convergence across variables. “Within chain” means ESS is calculated within each chain, but quantiles are across chains (i.e., quantiles of a list of (n chains) × (n parameters) values). “Across chain” ESS is calculated using relative variance within chain and across chains. The columns are: 1) only elements of contemporaneous coefficients matrix $A_0$; 2) only state variances; and 3) the impulse response of each variable to each shock at the 1 month, 12 month, and 48 month horizons.

$$PSRF = \sqrt{1 + \frac{m}{ESS}}$$, where $m$ is the number of MCMC chains. The PSRF can be interpreted as the scale by which the parameter distribution could be reduced if simulations were continued to infinite length. The reference Gelman et al. (2014) suggests a “safe” threshold value of $PSRF < 1.1$, which is mostly (but not completely) met by our impulse response parameters. In particular, for the $t$-distributed, full data model, 85% of impulse response parameters corresponding to shocks 3 (HH credit expansion), 6 (monetary policy), 9 (GZ spread), and 10 (interbank spread) are less than 1.1.

We consider these results sufficiently converged because the impulse response functions are ultimately the parameters of greatest interest for our analysis. Furthermore, pooling across several chains produces a conservative estimate of posterior variance, and the additional posterior variance that is attributed (in our sample) to cross-chain differences is, in the scale of economic interpretation and relative to the additional uncertainty in sampling the $(A_i)_{i \geq 1}$ coefficients, not so large.
<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$\Lambda$</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>1.13</td>
<td>1.44</td>
<td>1.01</td>
</tr>
<tr>
<td>50%</td>
<td>1.31</td>
<td>1.78</td>
<td>1.02</td>
</tr>
<tr>
<td>75%</td>
<td>1.68</td>
<td>2.66</td>
<td>1.04</td>
</tr>
<tr>
<td>&lt; 1.1 %</td>
<td>24%</td>
<td>4%</td>
<td>94%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$\Lambda$</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>1.05</td>
<td>1.14</td>
<td>1.00</td>
</tr>
<tr>
<td>50%</td>
<td>1.13</td>
<td>1.42</td>
<td>1.01</td>
</tr>
<tr>
<td>75%</td>
<td>1.29</td>
<td>2.12</td>
<td>1.02</td>
</tr>
<tr>
<td>&lt; 1.1 %</td>
<td>38%</td>
<td>20%</td>
<td>98%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$\Lambda$</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>1.07</td>
<td>1.27</td>
<td>1.01</td>
</tr>
<tr>
<td>50%</td>
<td>1.10</td>
<td>1.18</td>
<td>1.00</td>
</tr>
<tr>
<td>75%</td>
<td>1.22</td>
<td>1.41</td>
<td>1.00</td>
</tr>
<tr>
<td>&lt; 1.1 %</td>
<td>48%</td>
<td>36%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$\Lambda$</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>1.04</td>
<td>1.09</td>
<td>1.00</td>
</tr>
<tr>
<td>50%</td>
<td>1.10</td>
<td>1.18</td>
<td>1.00</td>
</tr>
<tr>
<td>75%</td>
<td>1.22</td>
<td>1.41</td>
<td>1.00</td>
</tr>
<tr>
<td>&lt; 1.1 %</td>
<td>48%</td>
<td>36%</td>
<td>100%</td>
</tr>
</tbody>
</table>

### Table 3. Distribution of potential scale factor reduction (PSRF) measures of posterior simulation convergence across variables. The columns are: 1) only elements of contemporaneous coefficients matrix $A_0$; 2) only state variances; and 3) the impulse response of each variable to each shock at the 1 month, 12 month, and 48 month horizons. Numbers are based on 6000 draws per chain subsampled from 60,000 (first three models) and 4000 draws subsampled from 40,000 ($t$ up to 12/07 model).

#### 3.4. Marginal Likelihood Estimates. We numerically estimate the marginal data density (MDD) of various models using the modified harmonic mean formula implemented by Geweke (1999). This approach uses the following identity for some probability density $g$ (with conditioning on $F$ omitted):

$$
\mathbb{E} \left[ \frac{g(\theta)}{P[Y \mid \theta]} \right] = \int \frac{g(\theta)}{P[Y \mid \theta]P[\theta]} P[\theta \mid Y] d\theta
= \left( \int P[Y \mid \theta'] d\theta' \right)^{-1}
$$

This is the inverse of the marginal data density that we are targeting. As a Monte-Carlo approximation to the above we take the following:

$$
\mathbb{E} \left[ \frac{g(\theta)}{P[Y \mid \theta]} \right] = G^{-1} \sum_{i=1}^{N} \frac{g(\theta^{(i)})}{P[Y \mid \theta^{(i)}]P[\theta^{(i)}]}
$$

where $i = \{1 \ldots N\}$ index individual draws.
In our specific case, we rely on the Monte Carlo method to numerically integrate over \( \tilde{\theta} := \{ \theta_1, \theta_3 \} = \{ A_0, \{ A_m \}_{m=1}^M, \{ \xi_{i,t} \}_{i,t=1}^n \} \), since we have access to the posterior density already integrated over \( \theta_2 \). We set \( g(\tilde{\theta}) = f(\theta_1)g(\theta_3) \), where \( f(\theta_1) \) is a Gaussian approximation of the distribution of \( \theta_1 \), truncated to \( \rho \% \) of its probability mass to control the effects of outliers, and \( g(\theta_3) \) is the appropriate prior distribution for the \( \{ \xi_{i,t} \}_{i,t=1}^n \). We make separate estimates for each of several MCMC chains, and use the comparison between chains to (very roughly) estimate the variance in our estimators (which, in our experiences, tends to be underestimated by what asymptotic theory in the number of draws and chains suggestions).

Table 4 gives some sense of the dispersion in our marginal data density estimates for \( \rho = 0.95 \). Results for other \( \rho \in [0.90, 0.99] \) were very similar (i.e., within 1 log point or less). In absolute terms, the consistency between chains is low (a 5 log point difference implies a substantial \( \exp(5) \approx 150 \) odds ratio between two models). But the precision is enough to distinguish among our models, in the sense that no rankings are flipped when we use the minimum estimate for the higher density model and the maximum estimate for the lower density model.

### Table 4. Marginal data density estimates for different models and different truncations of the Gaussian weighting distribution.

<table>
<thead>
<tr>
<th>Model</th>
<th>Chains</th>
<th>Median</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t ), full data</td>
<td>6</td>
<td>52490.13</td>
<td>52475.76</td>
<td>52503.58</td>
</tr>
<tr>
<td>multinomial, full data</td>
<td>3</td>
<td>52400.47</td>
<td>52400.41</td>
<td>52400.59</td>
</tr>
<tr>
<td>Gaussian, full data</td>
<td>4</td>
<td>51768.38</td>
<td>51761.44</td>
<td>51770.07</td>
</tr>
<tr>
<td>Gaussian, full data with 3-year growth</td>
<td>3</td>
<td>51225.91</td>
<td>51251.50</td>
<td>51260.05</td>
</tr>
<tr>
<td>( t ), data up to 12/07</td>
<td>6</td>
<td>48796.70</td>
<td>48768.81</td>
<td>48811.44</td>
</tr>
</tbody>
</table>

4. CONSTRUCTION OF CREDIT DATA

This section provides more details about our selection and treatment of aggregate credit data and a comparison with the quarterly frequency data available in the U.S. Flow of Funds. Later, in Appendix 8, we investigate how switching between these two data sources changes our main results.

For our main VAR analysis, we use total loans data from the Federal Reserve’s H.8 “Assets and Liabilities of Commercial Banks in the United States” dataset. These data are
compiled from a sample of approximately 875 domestically chartered banks and “foreign-related institutions” as of the most current release. These constitute, to our knowledge, the most complete monthly or higher frequency source of historical credit data in the United States. There are conceptual barriers to extrapolating conclusions with these data to conclusions about household credit and firm credit more properly defined. Our measure of “household” credit includes both household and commercial real estate, and our measure of “firm” credit includes only commercial and industrial loans. Additionally the subset of credit issued by commercial banks may not be representative of all credit in the economy. In spite of these obstacles, we argue that a commercial bank data series with some corrections for large “measurement breaks” (including bank mergers, accounting standards changes, and other “non-fundamental” sources of variation) closely matches the dynamics of more precisely defined, lower-frequency measures. Thus we expect (and find) that results would be very similar using a different, quarterly-frequency dataset.

4.1. **Correction for Measurement Breaks.** At several points in the raw credit data, we found big outliers in growth rates that corresponded to special notes in the H.8 data releases. Under the assumption that these observations do not represent relevant variation

---

7A full documentation can be found at [https://www.federalreserve.gov/releases/h8/About.htm](https://www.federalreserve.gov/releases/h8/About.htm).

8An alternative option we do not pursue is interpolating the quarterly data with these monthly estimates. The Federal Reserve’s G.19 “Consumer Credit” series, for instance, does something similar. We have (standard) concerns about endogeneity issues in time-series regressions.
in credit conditions, we manually “corrected” the data by adding or subtracting the quantities attributed to these events in the documentation (Table 5). Note that these changes affect the entire time series after the break date in terms of levels.

Figures 2 and 3 show these changes in log difference units for household and business credit measures respectively. It is worth noting that our correction does not completely smooth over the accounting change in April 2010, presumably because there is some change in balance sheets in anticipation in the previous period. We suspect that the model with normal mixture errors would not be very affected by one relatively small outlier, but an alternate approach that we have not yet implemented is to correct around a linear trend for that three month stretch.

In a model with Gaussian errors and heteroskedasticity, the pre-correction data series generated outliers up to 7 standard deviations and consequently had a significant effect on model fit (entirely removing our result of a weakly significant negative output response to a household credit shock).

4.2. Comparison with Quarterly Data. Many other studies have used the quarterly credit series published in the Federal Reserve’s Z.1 “Financial Accounts of the United States.” These data are separate for the balance sheets of “Households and nonprofit organizations,” “Nonfinancial corporate businesses,” “Nonfinancial noncorporate businesses,” “State and local governments,” the “Federal government”, “Domestic financial sectors,” and the “Rest of the world.” The Bank for International Settlements (BIS), in its multi-country panel dataset, reports versions of the “Households and nonprofit organizations” and “Nonfinancial corporate business” series adjusted for breaks.

Figure 4 compares measures of household credit in our dataset versus the single series in the quarterly dataset. In general, we find that the sum of consumer and real estate credit closely approximates the dynamics of the quarterly series, though perhaps with an extra large contraction after the 2008 financial crisis. A series constructed with just consumer credit (in blue), while it does avoid the issue of including real estate loans to businesses, 

---

9 The corrections for March and April relate to the same event, which occurred in the last week of March. Our monthly data are simple averages of these weekly data. So our solution was to make 1/4 of the correction in March (i.e., to correct for 1/4 of the averaged data which was affected) and 3/4 of the correction in April. Figures 2 and 3 show that the resulting data series look reasonably smooth.
Figure 2. Log differences of household credit series before (green) and after (blue) corrections. Vertical red lines denote the dates of changes.

Figure 3. Log differences of business credit series before (green) and after (blue) corrections. Vertical red lines denote the dates of changes.
seems to grow at a significantly slower rate and include different medium-frequency (3-4 year) cyclical patterns.

Figure 5 plots a similar comparison for the business credit series, but with multiple options at the quarterly frequency. The monthly data (in red) does not so clearly approximate the blue series of nonfinancial corporate credit, which is favored by the BIS (and other panel studies which use that dataset). It does, however, capture some of the same long-term trends (including accelerations and slow-downs around business cycles), and much more closely approximates the nonfinancial noncorporate series (in green) and the average between the two quarterly series (in orange).

5. LARGE RESIDUALS IN THE MAIN MODEL

Table 6 displays the 10 largest residuals unscaled by λ, on the left, and the 10 largest scaled (i.e. “surprise units”) residuals on the right. The two 9/11 residuals are more prominent as surprises.

Five of the ten biggest surprises are in 2007-8. This concentration in time suggests that our specification of variance regime periods could be improved.
**Figure 4.** Comparison of household credit series, in log levels normalized to zero at the beginning of 1973. The data series we use in our main analysis is in red.

**Figure 5.** Comparison of firm credit series, in log levels normalized to zero at the beginning of 1973. The data series we use in our main analysis is in red.
The far right column of the table shows that these residuals, though far out in the tails of the $t(5.7)$ distribution we have assumed for them, are not totally unexpected. However, with the approximately 5000 residual values from our model, the $t(5.7)$ distribution makes the expected number of surprise residuals as big as 7.52 only 1.8, whereas we have estimated 10 of this size. This discrepancy is easily big enough to reject the $t(5.7)$ distributional hypothesis, which suggests another direction in which our model specification could be improved.

6. ROBUSTNESS OF THE MAIN MODEL

Here we discuss small perturbations of the paper’s main VAR specification. These include changes to the distribution of structural shocks, the sample period, and the method of identification, the assumed linear specification of credit growth rates.

6.1. Different Error Distributions. Figures 6 and 7 present the full full impulse responses for the models with $t$ and normal error distributions, respectively. Most of the impulse responses are similar, and in most cases the error bands are wider for the normal model than for the $t$ model. This is what would be expected if the $t$ model were correct, since in that case the normal model would still give consistent estimates, but would be provide less efficient estimates.

There are a few notable differences between results from the models with Gaussian and $t$-distributed shocks. The differences between the two models can be seen in Figure 8, which compares the median impulse responses for IP, HHC, BC, R, PCM, GZ and ES to shocks 3 (HHC), 4 (BC), 6 (R, or monetary policy), 7 (PCM), 9 (GZ) and 10 (ES), with each shock scaled so that its largest initial component is the same size for both the $t$ and Gaussian models. The differences between them are mainly within the error bands, with four major exceptions. The normal errors model puts more posterior probability on a nonzero output effect for the household credit shock and the interbank spread shock. The output response of the former, with normal errors, is comfortably significant (less than zero) with 68% bands and barely significant with 90% bands (which are not pictured). The output response of the latter is comfortably significant, from the first 18 months of response, at both levels.
The Gaussian errors model shows a substantial negative response of output to shock 7 (PCM), whose largest component is a persistent increase in commodity prices. The shock also induces a modest, but significantly positive, upward movement in BC. Since this shock moves PCM immediately and BC with a delay, it is best interpreted as a commodity price shock, not an autonomous credit shock. It nonetheless implies a component of variation with a negative correlation between business credit and later output growth.

The normal model places considerably less posterior probability on a negative output response for a business credit shock, though in this case the $t$ model’s response is well within the normal model’s error bands.

Note that these models are not simply alternative views of the data. The $t$-distributed errors model fits better than the normal model by a wide margin. To the extent that the Gaussian and $t$ models differ, this is likely due to the Gaussian model being distorted by large outliers that are downweighted in the $t$ model.

We also experimented with a mixture-of-normal errors model, which is not discussed in the main paper and is included in this appendix mainly as a robustness check. It sets the structural residual $\varepsilon_{i,t} \sim N(0, \lambda_{i,t}\xi_{i,t})$ with $\xi_{i,t} = \beta_i$ with probability $\alpha_i$ for $i \in \{1 \ldots k\}$ (11)

We choose $k = 3$ and set $\beta_1 < \beta_2 < \beta_3$, which provides an intuition of “low, medium, and high” variance options for each observation. Based on maximum likelihood estimation with residuals from the normal-errors model, we set $\beta_i = \{.46, 1.37, 6.50\}$ and $\alpha_i = \{.59, .39, .02\}$. The normal-mixture model produced impulse responses and fit in between the Gaussian and $t$ models for the most part (Figure 9).

### 6.2. Shortened Sample Period.

Figure 10 shows the full impulse responses from the $t$ model estimated with data through 12/07. The impulse responses are quite similar to those from the full sample, with no differences that merit singling out. This suggests that the impact of spread shocks on macro variables is not special to the 2008-9 crash and was estimable from earlier data.
Figure 6. Impulse responses to the ten orthogonal structural shocks in the $t$ errors model over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions. Scaled to an “average” period with unit scale.
Figure 7. Impulse responses to the ten orthogonal structural shocks in the Gaussian errors model over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions. Scaled to an “average” period with unit variances.
Figure 8. Impulse responses for shocks 3, 4, 6, 7, 9 and 10, Gaussian model in blue, t model in red, scaled to have identical sized initial shocks in HHC, BC, R, PCM, GZ and ES, respectively. The solid lines are from the posterior mode and median respectively, and dotted lines are 68% bands.

6.3. Triangular Normalization. Figures 11 and 12 show the full impulse responses of two Gaussian-errors models with triangular restrictions on short-run responses. The former uses these restrictions for identification (i.e., is a standard “Cholesky” recursive VAR
FIGURE 9. Impulse responses to the ten orthogonal structural shocks in the model with discrete normal mixture distributed errors over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions. Scaled to an “average” period with unit variances.
FIGURE 10. Impulse responses to the ten orthogonal structural shocks in the $t$ errors model, estimated with data up to December 2007, over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions. Scaled to an “average” period with unit scale.
Figure 11. Impulse responses to 10 structural shocks in a Cholesky-identified VAR with constant structural variances over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions.
Figure 12. Impulse responses to 10 structural shocks in a model with triangular restrictions and heteroskedasticity over 60 months, with 68% (dark blue) and 90% (light blue) posterior uncertainty regions.
Table 7. Correlation among shocks (indexed by number, named in the columns) in two triangularized models, without regime heteroskedasticity (rows) and with regime heteroskedasticity (columns).

<table>
<thead>
<tr>
<th>Cholesky Model</th>
<th>Heteroskedasticity Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (BC)</td>
<td>0.996</td>
</tr>
<tr>
<td>6 (R/MP)</td>
<td>-0.008</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>0.005</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>-0.009</td>
</tr>
<tr>
<td>4 (BC)</td>
<td>0.159</td>
</tr>
<tr>
<td>6 (R/MP)</td>
<td>0.883</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>0.002</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>-0.003</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>0.148</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>0.135</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>-0.016</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>0.114</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>0.943</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>0.165</td>
</tr>
<tr>
<td>9 (GZ)</td>
<td>0.017</td>
</tr>
<tr>
<td>10 (ES)</td>
<td>0.916</td>
</tr>
</tbody>
</table>

with no heteroskedasticity). The latter includes heteroskedasticity and imposes short-run restrictions as over-identifying.

The Cholesky model includes a sizable long-term output contraction (about 0.4% over 48 months, explaining 13.0% of variance at that horizon) in response to a business loans shock. This shock, like the corresponding shock in the baseline (no restrictions, with heteroskedasticity) model, includes some monetary tightening in the response to this fourth shock. The modest tightening of monetary policy and credit spreads suggests that this shock at least partially represents business (but not household) borrowing in the midst of both monetary stringency and financial stress. On the other hand, the overidentified (triangular model with regime-switching variance) model produces impulse responses that look much like their counterparts in the unrestricted Gaussian-errors model.

Table 7 sheds some light on the discrepancy by reporting the correlations between shocks 4 (BC, business credit), 6 (R, monetary policy), 9 (GZ, corporate bond spread), and 10 (ES, inter-bank (Eurodollar) spread) in the two models. The business credit shock in the plain, Cholesky-identified VAR (rows) “absorbs” small amounts of each of the three interest rate shocks in the model with regime switching heteroskedasticity. Breaking this down by regime illustrates the pattern more closely. In the period of aggressive Federal Reserve action (October 1979 to December 1982), the Cholesky VAR’s business loans shock is very closely correlated with the other model’s monetary policy shock (0.467). In the financial crisis and Great Recession (January 2008 to December 2010), the Cholesky VAR’s business loans shock is positively correlated with the corporate bond and inter-bank shocks (0.644 and 0.620 respectively) and negatively correlated with the monetary policy shock (-0.206) to partially offset the effect. In this context, identification through
heteroskedasticity delivers a different, and perhaps \textit{a priori} more reasonable, separation of policy, financial stress, and credit expansion effects than a simple triangular orthogonalization.

6.4. \textbf{Long-term Growth Rates}. Our estimated system uses 10 months of lagged values, or with quarterly data four lags. Of course an autoregressive system with 10 lags can capture dynamics that operate over much longer spans of time than 10 months, but certain kinds of dynamics, e.g. pure delays of over 10 months, cannot be well approximated by a 10-lag autoregressive system. As a check, we try estimating our main (Gaussian errors) specification with three-year growth rates of household and business credit instead of levels. Figure 13 displays the impulse responses of output, prices, and the differenced credit aggregates in this model’s two “credit aggregate” shocks. The household credit shock in this model explains a large proportion of long-term forecast errors in the growth rate of credit (29% at 5 years), but still only a modest amount (6%) of GDP variation at the same horizon.

We can directly and easily compare the fit of this model and the original models through posterior odds conditional on initial conditions.\textsuperscript{10} The differences model fits considerably worse than the levels model, with an estimated marginal likelihood of 51225.91 (compared to 51768.38). So even if the model did imply a very strong connection between credit growth and recessions, we would have strong evidence to favor the original model.

6.5. \textbf{Non-linear Transformation}. Potentially our method is failing to capture a positive long-term relationship because it considers only linear effects. This implies that small credit movements have effects in proportion to large ones, and that negative shocks have an equal and opposite effect as positive shocks.\textsuperscript{11}

While extensive exploration of possible nonlinearities in the model would have to be a new research project, we did try applying a smooth non-linear transformation to the 3-year growth rates of credit that allowed increased weight on large positive growth rates. The idea was to explore the hypothesis, put forward in other research, that modest credit

\textsuperscript{10}The Jacobian of any linear transformation like this is one. But now, if the model is with three three-year growth rates, we are implicitly conditioning on the values of \{c_0 - c_{-36}, c_{-1} - c_{-37}, \ldots\}, where \textit{c} is the credit variable. This is different but not higher dimensional information.

\textsuperscript{11}Jordà, Schularick and Taylor (2013), for instance, focus on upper \textit{n} percentile credit growth events.
expansion has no negative effects, while unusually rapid credit expansion does create future problems. We considered transforming credit growth according to the function

\[ f(c_t) = \begin{cases} 
  c_t & \text{if } c_t < a \\
  \alpha_1 + \alpha_2 c_t + \alpha_3 c_t^2 & \text{if } a \leq c_t < b \\
  \beta c_t & \text{if } c_t \geq b
\end{cases} \]

for \( c_t \) as credit growth rates, \( b > a > 0 \), and with the coefficients \( \alpha_i \) and \( \beta \) calibrated to make slopes and levels continuous at \( a \) and \( b \). This allowed the “extra weight” \( f'(c_t) \) to get
larger without getting unboundedly big. We looked for a posterior mode optimizing over $a$, $b$, and $\beta$, as well as the other parameters of the model.\textsuperscript{12} In all optimization exercises to find the posterior mode of such a model, the data favored models without any nonlinear transformation.\textsuperscript{13}

7. Models with Fewer Variables

In this section, we estimate a VAR model with only the following variables: industrial production, the price level, and the two credit variables. It is possible to obtain results suggesting that credit shocks lead to long-term output contractions. We argue based on Monte Carlo simulations (in analogy with our analysis of single-equation models in Section V) that the small-system results are consistent with our main, larger model.

7.1. A Small VAR Model. Figure 14 presents the impulse response of the aforementioned four-variable model, identified with heteroskedasticity. In such a model, a shock associated a persistent increase in business loans is associated with a greater than half a percentage persistent reduction in output. This shock (at posterior mode parameters) accounts for 58\% of business loan variation and 7\% of output forecast error variation in one-quarter-ahead forecasts, and 22\% and 8\% respectively in five-year-ahead forecasts. This is considerable relative to the fraction of output variation produced by the corresponding shock in the 10-variable model.\textsuperscript{14}

If one estimates a four-variable VAR without regime-changing heteroskedasticity (and identifies structural shocks with Cholesky ordering), the results are qualitatively similar to those in Figure 14 but with tighter error bands.

7.2. Monte Carlo Exercise. We can shed some light on this by simulating data from the posterior mode point estimates of the large model, then estimating the same small models on the simulated data. Like in Section V, we simulated 6,000 samples from the posterior mode of the $t$-errors model. and, for each artificial time series, estimated a four-variable Cholesky-identified VAR (including output (IP), the price level (P), household

\textsuperscript{12}We used a coarse grid over $a$ and $b$ (based on matching quantiles of the observed distribution) and searched continuously over $\beta$, with an exponential prior with parameter 2.

\textsuperscript{13}In this specification, $\beta_1 = 1, \alpha_1 = \alpha_3 = 0, \alpha_2 = 1$.

\textsuperscript{14}In the larger model, two shocks together explain approximately the same first period variance. One is the 4th shock (which explains 38.8\% of variance at the posterior median) and the other is the 2nd shock (which explains 26.2\%). Neither produces a negative correlation between business credit and output.
We obtained negative and “large” (greater than a 0.5% output contraction) response to a one-standard deviation credit shock with the probabilities shown in Table 8. These calculations are done with data at different frequencies, aggregated by taking averages of the monthly data our model generates. Observing negative point estimates (suggesting a negative impulse response of output to credit shocks) is not so uncommon, particularly as the frequency of data is reduced to annual.

The posterior simulations show that pattern of results observed in the real data is not unlikely, with 11% of samples showing a strong negative effect of BC shocks on output and none showing an effect of HHC shocks on output. With more highly time aggregated data, as has been used in some previous studies, our model implies that the probability of finding BC shocks predict substantially negative future output growth is stronger.

\footnote{These are the point estimate impulse response from lower triangular (Cholesky) identified VARs, with “standard” coefficient shrinkage, unit root, and covariance priors.}
### Table 8

<table>
<thead>
<tr>
<th></th>
<th>HH Credit</th>
<th>Bus. Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly</td>
<td>0.21</td>
<td>0.75</td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.29</td>
<td>0.75</td>
</tr>
<tr>
<td>Annual</td>
<td>0.50</td>
<td>0.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HH Credit</th>
<th>Bus. Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.02</td>
<td>0.30</td>
</tr>
<tr>
<td>Annual</td>
<td>0.16</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 8. Probability, in simulated draws from the posterior distribution of the $t$-model for the data, of a negative or “economically significant” (more than 0.5% negative) 3-year response of IP to a positive credit shock in a 4-variable VAR. Based on 6,000 simulated data series. Parameter draws that implied explosive behavior were treated as not having $\hat{\beta} < 0$, so these proportions are lower bounds.

8. Models with Quarterly Data

Figure 15 displays the results from estimating our model with a normal error assumption on quarterly averaged data. With quarterly data, the monetary policy shock (MP) and the interbank spread shock (ES) are not well separated. The monthly data show interest rates responding with some delay, but within a couple of months, to the ES shock. The quarterly data show the interest rate response to the ES shock as larger and within-quarter. The monthly data show the interbank spread responding to the MP shock, but not as strongly as to the ES shock. In the quarterly data, the MP shock accounts for most of the variance of the interbank spread (ES) variable. Because the quarterly data seem to partially confound the monetary policy shock with the spread shock, we omit detailed discussion of the rest of the impulse responses.

Nonetheless it is worth noting that in Figure 16, where we have substituted the Flow of Funds versions of our credit aggregates for our own measures, there is evidence that the definition of the credit aggregates does matter. In the flow of funds results, the HHC shock moves the price level down by a statistically significant, but small, amount, and immediately lowers the term spread without much immediate effect on R — i.e. it lowers
FIGURE 15. Impulse responses with quarterly-averaged data. Heteroskedasticity with change dates as in the main monthly model. Normal errors assumption.
long rates. The shock is followed with a delay by negative output growth. With our data quarterly averaged, neither the HHC or BC shocks (which in the monthly data we have interpreted as “credit aggregate shocks”) is followed by substantial negative output growth, and the simultaneous jumps in P and TS in response to the HHC shock that appear in the flow of funds data results are not present.

These results make clear that the Flow of Funds aggregates and the data we are using behave somewhat differently. It is possible, though, to reconcile these results with some more qualitative aspects of our main model’s story. The credit shocks in the model estimated with Flow of Funds data explain only small fractions of credit and output variance. At the 20 quarter (5 year) horizon, this HHC shock explains only 2.7% of forecast error variance in HHC itself and only 7.6% of forecast error variance in IP. Moreover, both credit shocks (to HHC and BC) spike in relative variance during the variance regime of the extreme monetary policy (Q1 1980 to Q4 1982) rather than the Great Moderation, in build-up to the Great Recession (Table 9).
REFERENCES


