

ALGORITHM-DRIVEN SVARS: NAVIGATING THE WILDERNESS OF BIG DATA

YUCHENG YANG AND TAO ZHA

ABSTRACT. Every SVAR result is conditional on two choices: the restrictions that identify the shock and the variables on which they operate. The literature disciplines the first; the second is chosen by hand. We develop a Bayesian methodology that constructs information sets, uses an out-of-sample criterion, and retains the largest system it admits. Under recursive identification, output rises with housing production rather than household credit alone. For monetary policy, an anchor-free joint Bayesian proxy SVAR with multiple instruments strengthens the credit spread channel. A core system augmented with the selected corporate spread identifies expected default risk as a potent transmission margin.

Date: 4th August 2026.

Key words and phrases. Information set, model construction, out-of-sample selection, external instruments, household credit, housing production, expected default risk, monetary policy transmission.

JEL classification: C55, C53, E00, E52, E44, C32

Yang: University of Zurich and SFI; Zha: Federal Reserve Bank of Atlanta, Emory University, and NBER. We thank Lukas Hack, Tom Sargent, and participants at various seminars and conferences for helpful discussions, as well as Tong Xu for his initial involvement in this project. Hongyi Fu provided outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the National Bureau of Economic Research.

I. INTRODUCTION

Structural vector autoregressions (SVARs) produce empirical evidence, under identifying restrictions, about the dynamic effects of economic shocks. That evidence often becomes the object that theory aims to explain. Many models are built to match impulse responses estimated from SVARs (Christiano et al., 1999; Sims and Zha, 2006). Every SVAR result is conditional on two choices: the restrictions that identify the shock and the variables on which those restrictions operate. The first choice is formalized, defended, debated, and, most important, disciplined. The second is usually left to the researcher. The information set behind the impulse responses is therefore not preliminary housekeeping. It is part of the evidence.

Under standard practice, a researcher begins with an economic question, chooses a set of variables by hand, imposes identifying restrictions on the resulting system, and studies the impulse responses. This practice was understandable when data were limited and computation was costly. It is no longer compelling when researchers observe hundreds of macroeconomic, financial, credit, housing, labor market, price, and interest rate series. A hand-built information set is not a neutral starting point. It is an undisciplined model selection decision. Omitting relevant variables can alter the innovations to which identification is applied and the impulse responses that theory is subsequently asked to explain. In the wilderness of big data, hand selection is less credible across researchers, while estimation of a system containing every possible candidate is impractical.

This paper develops an algorithm-driven Bayesian methodology for tackling this issue head-on by constructing the information set of an SVAR. The methodology rests on a clear division of labor. The researcher chooses the economic question, the core variables that define it, and the identifying restrictions. The algorithm constructs the information set. Bayesian out-of-sample (OOS) evidence selects its complexity. The method is not an unsupervised search that asks the data to choose the question or the identification. A study of household credit must begin with household credit and the outcomes the question is meant to address. A study of monetary policy must begin with the policy and transmission variables needed for that question. The data discipline the surrounding information set only after the economic question and its objects have been specified.

The methodology consists of two components. First, the model construction procedure searches over variables and factors outside the core that contain predictive information for the composite disturbances of the current system. It is an iterative procedure. At each iteration, the current SVAR is estimated at its posterior mode and its fitted composite disturbances are computed. An auxiliary optimization step determines whether the contemporaneous or lagged values of variables outside the current system contain information about those disturbances for each value of a model complexity parameter. Selected variables enter the system, the enlarged SVAR is reestimated, and the disturbances are updated. The procedure

stops when no remaining variable enters and then delivers a terminal information set for that value of the model complexity parameter. Across the prespecified values of model complexity, it produces terminal systems of different dimensions and does not favor a small or a large model in advance.

Second, the Bayesian OOS criterion determines which complexity to retain. A validation sample that plays no role in model construction is used to evaluate forecasts of the core variables. For each loss, the rule first retains terminal systems that are credibly better than the reference. If none exists, it retains the reference and systems whose loss differences are not credibly distinguishable from zero. It then selects the largest system in the union of the squared and absolute loss comparison sets. The objective is information retention, not mechanical sparsity. The output is an interpretable SVAR with the selected information set for a specified economic question. The methodology produces an algorithm that is reproducible and fast.

We provide the theoretical foundation for this methodology. For every value of the model complexity parameter, the construction path is monotone, terminates after finitely many updates, and produces a unique terminal system. We derive an exact stopping characterization, establish that admission does not depend on measurement units, and show that sufficiently small numerical perturbations do not alter the terminal system (computationally stable). The Bayesian OOS step returns a unique model complexity parameter and hence a unique selected system. Together, these results turn specification search from hidden researcher choices into an open, reproducible algorithm and bring the construction of the information set under the same discipline long applied to identification.

This evolution matters not merely in theory but in practice. We apply the methodology to two of the most studied questions in the SVAR literature: the effects of household credit shocks and the effects of monetary policy shocks. Together, the two applications demonstrate its generality across two classic identification methods: recursive identification and proxy identification.

The first application relates to [Mian et al. \(2017, MSV hereafter\)](#), who find that output rises and later declines after a household credit shock. [Brunnermeier et al. \(2021, BPSS hereafter\)](#) replicate the same initial output boom in a larger monthly system and show that its interpretation depends on the information set. Both systems, however, are chosen by the researchers, and neither contains direct measures of housing production despite the central role of residential construction emphasized by [Leamer \(2007, 2015\)](#). Using MSV’s three variables—output, business credit, and household credit—as the core, our procedure selects a 13-variable system that contains four measures of housing production. No particular housing block is imposed in advance; each housing series competes with every other candidate on the same terms in model construction.

The selected information set changes the central economic conclusion, not merely its precision. A household credit shock raises household credit, but the posterior median of output

shows no meaningful initial boom and all four measures of housing production decline. A separate housing production shock raises permits, starts, industrial production, and household credit together. Within the selected system, output rises when housing production expands and does not rise when household credit expands while housing production contracts. This pattern holds under both the recursive identification of MSV and the heteroskedasticity identification of BPSS. It distinguishes household credit from housing production without assigning a primitive cause to the change from the smaller systems.

The second application develops an independent Bayesian method for proxy SVARs. A common two-stage implementation chooses one reduced-form innovation as an anchor and measures every response relative to it. [Gertler and Karadi \(2015, GK hereafter\)](#), for example, use the innovation in the one-year Treasury rate and one external instrument. With one instrument, admissible anchors recover the same normalized impact vector. With several instruments, however, the estimated responses can depend on the anchor in finite samples, and no economic principle selects one reduced-form innovation over another. The three instruments of [Swanson \(2021\)](#) capture distinct dimensions of monetary policy surprises—the federal funds rate, forward guidance, and large-scale asset purchases. Their joint deployment therefore provides no economic basis for privileging the reduced-form innovation of any single policy indicator as the anchor.

We establish that a vector of external instruments identifies one shock and its equation in a joint SVAR if and only if at least one instrument is correlated with that shock. The method allows any number of instruments for one shock, extends to several shocks and their corresponding instrument vectors, delivers explicit formulas for the identified equation and impulse responses, and estimates the proxy relation and the SVAR under one posterior distribution without a reduced-form innovation anchor. The model construction procedure and the joint proxy SVAR share a common principle. The first replaces a hand-chosen expansion of the information set; the second replaces a hand-chosen anchor. Both turn a consequential modeling choice into a transparent Bayesian procedure.

The empirical part of the second application studies monetary policy transmission. We begin with GK’s six core variables and use the three high-frequency instruments of [Swanson \(2021\)](#) jointly in our anchor-free proxy SVAR. The algorithm selects a 19-variable system. Rather than a dense ladder of closely related Treasury rates, it selects distinct labor market, housing, external, commodity, equity valuation, volatility, credit, and liquidity margins.

In the core system augmented with the selected GZ spread, the difference between the corporate bond spread and the excess bond premium of [Gilchrist and Zakrajšek \(2012, GZ hereafter\)](#) rises, with both the 68% and 90% credible bands above zero over the early horizons. Monetary policy tightening therefore affects not only the excess bond premium but also the expected default risk in this reporting system. Joint estimation with multiple instruments and no anchor strengthens GK’s central credit spread channel; the selected

corporate spread extends the evidence to expected default risk as an integral part of monetary policy transmission.

The remainder of the paper is organized as follows. Section II relates our contributions to the literature. Sections III–V develop the model construction procedure, establish its theoretical foundations, and prove the existence and uniqueness of the Bayesian OOS-selected system. Section VI presents the household credit application. Section VII establishes a necessary and sufficient identification theorem for proxy SVARs with multiple external instruments and applies the joint Bayesian method to monetary policy transmission. Section VIII concludes.

II. RELATED LITERATURE

Leamer (1978) provides the conceptual foundation for treating model selection and specification search as part of econometric inference rather than as hidden preliminary choices. Gu et al. (2020) provide a modern implementation of that principle by using separate training and validation samples to select model complexity in high dimensional predictive models. Our methodology brings this discipline into Bayesian SVAR analysis. The construction sample generates terminal information sets across values of model complexity, while a separate validation sample selects among them using posterior distributions of OOS forecast loss differences. We follow the posterior reporting convention of Sims and Zha (1999) and the sample splitting logic of Gu et al. (2020), rather than the expanding-window design of a real-time forecasting exercise such as Giannone et al. (2021).

The paper is about model selection, not model averaging. Most SVAR evidence used by economic theory comes from one selected system. An SVAR requires observed economic variables, a concrete information set, and identifying restrictions that deliver clear economic interpretations. Model averaging can be useful, but it answers a different question and can obscure the economic content of the system. Our objective is to select one SVAR for a given economic question. The procedure is designed for repeated empirical use rather than as a single computational exercise. Tractability is therefore part of the methodology, not a secondary convenience. A useful construction method must be fast, reproducible, and feasible for other researchers to replicate, extend, and apply to new economic questions.

Giannone and Reichlin (2006) show that omitted information can make shocks recovered from an SVAR nonfundamental. Forni and Gambetti (2014) build on this insight by using estimated factors to test whether a VAR is informationally sufficient and to amend it when deficiency is detected. Jarociński and Maćkowiak (2017) use posterior probabilities of Granger causal priority to rank and select variables for a Bayesian VAR. This strand of literature establishes that the information set matters. Big data poses a distinct model construction problem: how to build a question-specific SVAR from hundreds of observed candidate variables. Existing approaches have not tackled this joint construction and selection problem.

Factor methods provide a different response to big data. A factor augmented VAR compresses a large panel into a small number of latent factors and appends them to a VAR (Bernanke et al., 2005). In unpublished Walras–Bowley Lecture slides, Reichlin (2026) places factor-augmented vector autoregressions (FAVARs) and large Bayesian VARs within the same dense regularization principle. A FAVAR imposes a discrete cutoff at the selected factors, whereas a large Bayesian VAR continuously shrinks all directions and retains information beyond the factor space. The retrospective analysis highlights the empirical success of Bayesian VARs estimated in levels under Minnesota priors and the central role of OOS evaluation in selecting model complexity.

Our methodology follows the Bayesian VAR route but extends it from estimation to model construction. We retain the core variables that define the original economic question, use Bayesian shrinkage to estimate every expanded candidate system, and use OOS evidence to select its complexity. Unlike a FAVAR, the selected object remains a system of observed economic variables whose impulse responses have direct economic meaning. Unlike a full large Bayesian VAR, the method does not require every candidate variable to enter one enormous system. It constructs an information set tailored to the economic question, including persistent and nonstationary series with the data in log levels. Constructed factors in the FAVAR literature may enter the candidate set, but the methodology does not require the information set to be represented by a handful of latent factors.

Our anchor-free joint Bayesian proxy SVAR contributes to the external instrument literature. Mertens and Ravn (2013) formulate covariance restrictions that use proxies correlated with the shocks of interest and orthogonal to other shocks. Stock and Watson (2018) unify the SVAR and local projection approaches with instrumental variables (IVs) and clarify their identification conditions. GK implement the SVAR-IV approach by using one instrument and the reduced-form innovation in one policy indicator as an anchor. Arias et al. (2021) develop Bayesian proxy SVARs in a framework that pairs a k -dimensional proxy vector with a k -dimensional shock vector through a nonsingular $k \times k$ proxy–shock covariance matrix and thus does not accommodate multiple proxies for one shock. Their proxy equations may contain contemporaneous and lagged endogenous variables, which makes the exogeneity restrictions nonlinear functions of the structural parameters. Their importance sampler is not designed for the repeated posterior mode calculations required at every iteration and every value of the model complexity parameter. We instead keep the multiple instruments of Swanson (2021), constructed from high-frequency surprises, outside the SVAR system, consistent with GK and the earlier literature, and impose exogeneity directly on the covariance matrix.

Caldara and Herbst (2019) develop the closest Bayesian predecessor to our joint approach. Their published paper develops a model with one proxy for one shock. In that scalar case, their unrestricted measurement error variance is equivalent to our specification. Their published paper does not develop several proxies for one shock. Their online appendix extends the model to two proxies but assumes independent measurement errors. Applied to our

three-instrument setting, the same assumption would force every pair of measurement errors to be uncorrelated. Identification does not require this restriction. Our formulation allows an arbitrary number of instruments and leaves their covariance unrestricted, subject only to the proxy conditions required for identification. It therefore nests both their published one-proxy model and their independent-error extension.

Our contribution establishes the identification theory for this joint approach and generalizes it. We establish a necessary and sufficient population condition under which an instrument vector of arbitrary dimension identifies one shock and its equation, derive explicit formulas for the identified equation and impulse responses, and prove that the remaining equations need not be identified. We further establish population equivalence with the GK two-stage procedure, characterize their finite-sample difference with several instruments, and extend the result to several shocks with corresponding instrument vectors. Moreover, our augmented structural representation permits direct application of the efficient Gibbs sampler of [Waggoner and Zha \(2003\)](#), which makes the repeated high-dimensional proxy-SVAR estimation required by model construction computationally feasible.

The household credit application connects our results to the literature on housing production, credit allocation, and the business cycle. [Leamer \(2007, 2015\)](#) emphasize the central role of housing production in the business cycle. [Davis and Heathcote \(2005\)](#) document that residential investment leads GDP and is more volatile than nonresidential investment. [Rognlie et al. \(2018\)](#) show how overbuilding durable capital, especially housing, can create an investment hangover and a recession. [Müller and Verner \(2024\)](#) show that credit booms disproportionately finance households, construction, real estate, and other nontradable sectors, and that nontradable credit expansions predict subsequent growth slowdowns. Together, these studies establish the economic importance of distinguishing housing production from household credit when interpreting the output boom. Our OOS-selected SVAR makes this distinction: output rises when housing production expands, not when household credit rises alone.

The monetary policy application connects the expected default response to theory and firm-level evidence. In [Gomes et al. \(2016\)](#), a monetary policy tightening lowers inflation, raises the real burden of long maturity nominal debt, and increases corporate leverage and default. [Palazzo and Yamarthy \(2022\)](#) find that contractionary monetary surprises raise both the expected loss and risk premium components of firm credit default swap spreads, with larger responses among riskier firms. Our aggregate evidence complements these results. In the core system augmented with the selected GZ spread, expected default risk rises after a monetary policy tightening, with both the 68% and 90% credible bands above zero over the early horizons.

III. MODEL CONSTRUCTION

Let y_t be an $n \times 1$ vector of observed variables for $t = 1, \dots, T$. The data generating process is an SVAR with $\mathcal{L}$ lags,

$$y_t' A_0 = c' + \sum_{\ell=1}^{\mathcal{L}} y_{t-\ell}' A_\ell + \varepsilon_t', \quad (1)$$

where A_0 is an invertible $n \times n$ matrix of contemporaneous coefficients, A_ℓ is an $n \times n$ lag matrix for $\ell = 1, \dots, \mathcal{L}$, c is an $n \times 1$ vector of intercepts, and $\varepsilon_t \sim \mathcal{N}(0, I_n)$ is serially independent. The normalization $\text{Var}(\varepsilon_t) = I_n$ fixes the scale of the shocks. The identifying restrictions are imposed on A_0 and, when specified, on the lag matrices.

The vector y_t contains the full set of observed candidate variables; thus, its dimension n is large. For a given application, the analysis begins with three elements: a specific economic question, a set of core variables needed to address that question, and a set of identifying restrictions. The unresolved problem is which additional variables from the large candidate set should enter the SVAR information set. Selecting a small system by hand risks excluding relevant information, whereas estimating the entire candidate system is infeasible at the dimensions of big data. The obstacle is computational rather than conceptual.¹ We use the prior of [Sims and Zha \(1998\)](#), the Sims–Zha prior hereafter) to regularize the lag dynamics of each fixed system, but the prior does not determine which candidate variables belong in the model’s information set.

We formulate the choice of additional variables through the core equations. These equations address the original economic question. Let $y_t^\dagger$ contain the core variables and a subset of additional variables, and let $\tilde{y}_t$ contain the remaining candidate variables. To formalize the role of the remaining variables, partition

$$y_t = ((y_t^\dagger)', \tilde{y}_t')', \quad c = (c_1', \tilde{c}')', \quad \varepsilon_t = ((\varepsilon_t^\dagger)', \tilde{\varepsilon}_t')'.$$

In this partition, $\varepsilon_t^\dagger$ is the first block of the full-system shock vector ε_t . Conformably partitioning each coefficient matrix, Equation (1) becomes

$$\begin{aligned} & \begin{bmatrix} (y_t^\dagger)' & \tilde{y}_t' \end{bmatrix} \begin{bmatrix} A_{0,11} & A_{0,12} \\ A_{0,21} & A_{0,22} \end{bmatrix} \\ &= \begin{bmatrix} c_1' & \tilde{c}' \end{bmatrix} + \sum_{\ell=1}^{\mathcal{L}} \begin{bmatrix} (y_{t-\ell}^\dagger)' & \tilde{y}_{t-\ell}' \end{bmatrix} \begin{bmatrix} A_{\ell,11} & A_{\ell,12} \\ A_{\ell,21} & A_{\ell,22} \end{bmatrix} + \begin{bmatrix} (\varepsilon_t^\dagger)' & \tilde{\varepsilon}_t' \end{bmatrix}. \end{aligned} \quad (2)$$

¹A VAR with n variables and $\mathcal{L}$ lags contains $O(n^2 \mathcal{L})$ lag coefficients; posterior simulation for a large system is costly even in reduced form. Under a nonconjugate shrinkage prior of the Minnesota type, the key matrix factorization can expand from a $K \times K$ system to a full $Kn \times Kn$ system, where K is the number of predictors per equation. The resulting cost per MCMC draw is $O((Kn)^3)$.

The first block of (2) can be written as

$$(y_t^\dagger)' A_{0,11} = c_1' + \sum_{\ell=1}^{\mathcal{L}} (y_{t-\ell}^\dagger)' A_{\ell,11} + (u_t^\dagger)', \quad (3)$$

where

$$u_t^\dagger = \varepsilon_t^\dagger + r_t$$

and

$$r_t' = -\tilde{y}_t' A_{0,21} + \sum_{\ell=1}^{\mathcal{L}} \tilde{y}_{t-\ell}' A_{\ell,21}.$$

The composite disturbance vector $u_t^\dagger$ combines the first block of the full-system shocks with the contemporaneous and lagged contributions of variables outside the current system. This block representation provides the basis for model construction. The objective is to construct a system for $y_t^\dagger$ in which the remaining variables do not enter its equations:

$$A_{0,21} = 0 \quad \text{and} \quad A_{\ell,21} = 0 \quad \text{for every } \ell = 1, \dots, \mathcal{L}. \quad (4)$$

When (4) does not hold, contemporaneous or lagged values of variables outside the current system enter $u_t^\dagger$ through r_t . When it holds, $r_t = 0$, and hence $u_t^\dagger = \varepsilon_t^\dagger$. The decomposition yields a model construction criterion: the algorithmic procedure evaluates whether contemporaneous or lagged values of the remaining candidates contain predictive information for the fitted composite disturbance vector within the system of core variables.

The algorithm applies this criterion iteratively to construct the model. At each iteration, the current system is estimated and the fitted composite disturbance vector within the system of core variables is constructed. The model construction step evaluates whether contemporaneous or lagged values of the remaining candidates contain predictive information for at least one component of this vector. It admits the candidates chosen by the criterion and reestimates the enlarged system. Reestimation is essential because adding variables alters the fitted equations and the composite disturbance vector used at the next iteration.

Let

$$\mathcal{N} = \{1, \dots, n\}$$

index the full candidate set. We begin with an initial set $\mathcal{J}^{(0)} \subset \mathcal{N}$ of $n^{(0)}$ core variables chosen by researchers to address a particular economic question and taken as given by the model construction procedure. Let $y_t^{(0)}$ denote the corresponding subvector and set $k = 0$. At iteration k , let $\mathcal{J}^{(k)}$ denote the current index set, let $y_t^{(k)}$ denote the corresponding subvector, and define

$$n^{(k)} = |\mathcal{J}^{(k)}|.$$

The complement

$$\widetilde{\mathcal{J}}^{(k)} = \mathcal{N} \setminus \mathcal{J}^{(k)}$$

indexes the remaining candidates. Write $\tilde{y}_t^{(k)}$ for the corresponding $\tilde{n}^{(k)} \times 1$ vector, where $\tilde{n}^{(k)} = n - n^{(k)}$. For $j = 1, \dots, \tilde{n}^{(k)}$, let $\iota_j^{(k)} \in \mathcal{N}$ denote the index in y_t of the j th variable in $\tilde{y}_t^{(k)}$, such that

$$\tilde{y}_{j,t}^{(k)} = y_{\iota_j^{(k)},t}.$$

The distinction between j and $\iota_j^{(k)}$ is needed because the set of remaining candidates changes after each update.

At iteration k , reorder y_t and the corresponding rows and columns of each coefficient matrix so that the variables indexed by $\mathcal{J}^{(k)}$ appear first. For $\ell = 0, \dots, \mathcal{L}$, let

$$A_\ell = \begin{bmatrix} A_{\ell,11}^{(k)} & A_{\ell,12}^{(k)} \\ A_{\ell,21}^{(k)} & A_{\ell,22}^{(k)} \end{bmatrix}$$

denote the corresponding partition, and let $c_1^{(k)}$ denote the first block of the reordered intercept vector. Define $A_\ell^{(k)} = A_{\ell,11}^{(k)}$, $c^{(k)} = c_1^{(k)}$, and

$$(r_t^{(k)})' = -(\tilde{y}_t^{(k)})' A_{0,21}^{(k)} + \sum_{\ell=1}^{\mathcal{L}} (\tilde{y}_{t-\ell}^{(k)})' A_{\ell,21}^{(k)}. \quad (5)$$

Let $\varepsilon_{\mathcal{J}^{(k)},t}$ denote the subvector of ε_t ordered in the same manner as $y_t^{(k)}$. The composite disturbance associated with the current system is

$$u_t^{(k)} = \varepsilon_{\mathcal{J}^{(k)},t} + r_t^{(k)}. \quad (6)$$

The current system is therefore written as

$$(y_t^{(k)})' A_0^{(k)} = (c^{(k)})' + \sum_{\ell=1}^{\mathcal{L}} (y_{t-\ell}^{(k)})' A_\ell^{(k)} + (u_t^{(k)})', \quad (7)$$

subject to the maintained identifying restrictions. Equation (7) provides the exact block representation underlying the model-construction procedure. The algorithm operationalizes this representation by estimating the current finite-order SVAR and screening the remaining candidates against its fitted composite disturbances.

Collect the parameters of the current system in

$$\theta^{(k)} = (c^{(k)}, A_0^{(k)}, A_1^{(k)}, \dots, A_{\mathcal{L}}^{(k)}),$$

let Θ_k be the restricted parameter space, and write $Y^{(k)} = \{y_t^{(k)}\}_{t=1}^T$. The posterior mode is

$$\hat{\theta}^{(k)} \in \arg \max_{\theta \in \Theta_k} \{ \log p(Y^{(k)} \mid \theta) + \log \pi(\theta) \}, \quad (8)$$

where $p(Y^{(k)} \mid \theta)$ denotes the likelihood for θ based on $Y^{(k)}$, and $\pi(\theta)$ denotes the prior density appropriate for the dimension of the current system. If the maximizer is not unique, a prespecified deterministic rule selects one element of the argmax set.

Evaluating the fitted equations at the posterior mode gives

$$(\hat{u}_t^{(k)})' = (y_t^{(k)})' \hat{A}_0^{(k)} - (\hat{c}^{(k)})' - \sum_{\ell=1}^{\mathcal{L}} (y_{t-\ell}^{(k)})' \hat{A}_\ell^{(k)}, \quad t = \mathcal{L} + 1, \dots, T. \quad (9)$$

Evaluating the current-system equations at the posterior mode yields the fitted composite disturbance vector $\hat{u}_t^{(k)}$. By (6), the underlying composite disturbance combines the shocks associated with the current equations and the contribution $r_t^{(k)}$ from variables outside the current system. The model construction step screens the remaining candidates against these fitted composite disturbances.

For the model construction step, each variable in y_t is standardized before the first iteration using moments computed from the full data set.² For $i = 1, \dots, n$, define

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{i,t}, \quad \hat{\sigma}_i = \left[\frac{1}{T} \sum_{t=1}^T (y_{i,t} - \bar{y}_i)^2 \right]^{1/2},$$

and

$$y_{i,t}^{\text{std}} = \frac{y_{i,t} - \bar{y}_i}{\hat{\sigma}_i}.$$

We maintain $\hat{\sigma}_i > 0$ for every candidate variable. The values $\bar{y}_i$ and $\hat{\sigma}_i$ are computed once and held fixed throughout the model construction procedure. For $j = 1, \dots, \tilde{n}^{(k)}$, define the standardized remaining candidates by

$$\tilde{y}_{j,t,\text{std}}^{(k)} = y_{l_j^{(k)},t}^{\text{std}},$$

and stack their contemporaneous values and lags as

$$x_{t,\text{std}}^{(k)} = ((\tilde{y}_{t,\text{std}}^{(k)})', (\tilde{y}_{t-1,\text{std}}^{(k)})', \dots, (\tilde{y}_{t-\mathcal{L},\text{std}}^{(k)})')'.$$

This standardization applies only to the variables used in the auxiliary optimization and does not alter the variables used to estimate the current system.

The optimization procedure targets a fixed subset $\mathcal{Q} = \{q_1, \dots, q_{n^f}\} \subseteq \{1, \dots, n^{(0)}\}$ of the core equations, where $n^f = |\mathcal{Q}|$. Let $\hat{u}_{\mathcal{Q},t}^{(k)} = (\hat{u}_{q_1,t}^{(k)}, \dots, \hat{u}_{q_{n^f},t}^{(k)})'$ denote the corresponding subvector of fitted composite disturbances. To evaluate the remaining candidates, relate these disturbances to their standardized contemporaneous and lagged values:

$$(\hat{u}_{\mathcal{Q},t}^{(k)})' = (d^{(k)})' + (\tilde{y}_{t,\text{std}}^{(k)})' D_0^{(k)} + \sum_{\ell=1}^{\mathcal{L}} (\tilde{y}_{t-\ell,\text{std}}^{(k)})' D_\ell^{(k)} + (\eta_t^{(k)})', \quad t = \mathcal{L} + 1, \dots, T, \quad (10)$$

where $d^{(k)}$ and $\eta_t^{(k)}$ are $n^f \times 1$ vectors, and each $D_\ell^{(k)}$ is a $\tilde{n}^{(k)} \times n^f$ matrix. The vector $\eta_t^{(k)}$ is the residual in the auxiliary regression. Rows of $D_\ell^{(k)}$ index the remaining candidates, and columns index the equations in $\mathcal{Q}$. Equation (10) is used only for model construction. Its

²Section VI.1 describes the transformations applied to the raw economic variables before standardization. Each series $y_{i,t}$ is measured either in log levels or in percentage points.

coefficients measure whether the remaining candidates have predictive content for the fitted composite disturbances.

Let $T_{\mathcal{L}} = T - \mathcal{L}$. For equation $q \in \mathcal{Q}$, let $d_q \in \mathbb{R}$ and let

$$\beta_q = (\beta'_{0,q}, \beta'_{1,q}, \dots, \beta'_{\mathcal{L},q})' \in \mathbb{R}^{(\mathcal{L}+1)\tilde{n}^{(k)}}, \quad \beta_{\ell,q} \in \mathbb{R}^{\tilde{n}^{(k)}}.$$

Solve

$$\min_{d_q, \beta_q} \left\{ \frac{1}{2T_{\mathcal{L}}} \sum_{t=\mathcal{L}+1}^T [\hat{u}_{q,t}^{(k)} - d_q - (x_{t,\text{std}}^{(k)})' \beta_q]^2 + \lambda \|\beta_q\|_1 \right\}, \quad (11)$$

where the intercept d_q is not included in the ℓ_1 term and $\lambda > 0$ is the model complexity parameter selected by the out of sample (OOS) procedure described in Section V.³

Let $(\hat{d}_q^{(k)}(\lambda), \hat{\beta}_q^{(k)}(\lambda))$ denote a minimizer of (11), and partition

$$\hat{\beta}_q^{(k)}(\lambda) = ((\hat{\beta}_{0,q}^{(k)}(\lambda))', (\hat{\beta}_{1,q}^{(k)}(\lambda))', \dots, (\hat{\beta}_{\mathcal{L},q}^{(k)}(\lambda))')'.$$

For $\ell = 0, \dots, \mathcal{L}$, define

$$\hat{D}_{\ell}^{(k)}(\lambda) = [\hat{\beta}_{\ell,q_1}^{(k)}(\lambda) \quad \hat{\beta}_{\ell,q_2}^{(k)}(\lambda) \quad \dots \quad \hat{\beta}_{\ell,q_{n^f}}^{(k)}(\lambda)]. \quad (12)$$

Thus, $\hat{D}_{\ell}^{(k)}(\lambda)$ is a $\tilde{n}^{(k)} \times n^f$ matrix whose j th column contains the coefficients for equation q_j at lag ℓ , for $j = 1, \dots, n^f$.

The optimization in (11) is used only to construct the information set. For $j = 1, \dots, \tilde{n}^{(k)}$, collect the coefficients associated with the j th variable in $\tilde{y}_t^{(k)}$ across all equations in $\mathcal{Q}$ and across its contemporaneous and lagged values:

$$\hat{G}_j^{(k)}(\lambda) = [\hat{D}_{0,j}^{(k)}(\lambda), \hat{D}_{1,j}^{(k)}(\lambda), \dots, \hat{D}_{\mathcal{L},j}^{(k)}(\lambda)]. \quad (13)$$

Candidate $\iota_j^{(k)}$ is admitted when at least one coefficient in $\hat{G}_j^{(k)}(\lambda)$ is nonzero. Thus,

$$\mathcal{F}^{(k)}(\lambda) = \left\{ \iota_j^{(k)} : j = 1, \dots, \tilde{n}^{(k)}, \left\| \hat{G}_j^{(k)}(\lambda) \right\|_F > 0 \right\}. \quad (14)$$

The criterion in (14) admits a remaining candidate when its contemporaneous value or any of its lags has predictive content for at least one component of the fitted composite disturbance. Once admitted, the variable remains in the system at all subsequent iterations.

After solving (11), update the current index set according to

$$\mathcal{J}^{(k+1)} = \mathcal{J}^{(k)} \cup \mathcal{F}^{(k)}(\lambda). \quad (15)$$

If $\mathcal{F}^{(k)}(\lambda) \neq \emptyset$, form the enlarged subvector $y_t^{(k+1)}$, estimate the enlarged system, and compute $\hat{u}_t^{(k+1)}$ from (9). Reestimation is essential because admitting a variable alters both the

³The model-construction procedure incorporates machine learning tools used in modern empirical work, including the Lasso, into an auxiliary screening step. For each value of λ , the iterative procedure generates a terminal candidate information set. The dependent variables in the auxiliary optimization are the fitted composite disturbances evaluated at the posterior mode, and Bayesian posterior simulation evaluates the resulting terminal systems through OOS forecast loss distributions and selects the model complexity parameter λ .

fitted dynamics and the fitted composite disturbances against which the remaining candidates are evaluated.

For a fixed model complexity parameter λ , the construction terminates at iteration $\bar{k}(\lambda)$ when

$$\widetilde{\mathcal{S}}^{(\bar{k}(\lambda))}(\lambda) = \emptyset.$$

No remaining candidate is admitted for this value of λ . Denote the terminal index set, variable vector, fitted composite disturbance vector, and dimension by

$$\mathcal{J}^\dagger(\lambda) = \mathcal{J}^{(\bar{k}(\lambda))}, \quad y_t^\dagger(\lambda) = y_t^{(\bar{k}(\lambda))}, \quad \widehat{u}_t^\dagger(\lambda) = \widehat{u}_t^{(\bar{k}(\lambda))}, \quad n^\dagger(\lambda) = |\mathcal{J}^\dagger(\lambda)|.$$

The variables in $y_t^\dagger(\lambda)$ comprise the core variables and the additional variables selected at the model complexity parameter λ , and together form the terminal information set associated with λ . Section IV establishes finite termination and uniqueness of the mapping $\lambda \mapsto y_t^\dagger(\lambda)$.

To facilitate implementation, the complete procedure is summarized as follows:

- (1) **Initialize.** Begin with the core index set $\mathcal{J}^{(0)}$ associated with the economic application, fix a model complexity parameter λ from the prespecified grid, and set $k = 0$.
- (2) **Estimate.** Estimate the current system (7) under the prior π and compute the posterior mode $\widehat{\theta}^{(k)}$ in (8).
- (3) **Construct the fitted composite disturbances.** Compute $\widehat{u}_t^{(k)}$ from (9).
- (4) **Evaluate the remaining candidates.** Solve the auxiliary optimization problem (11) for the current system and construct $\widetilde{\mathcal{S}}^{(k)}(\lambda)$ from (14).
- (5) **Update.** Apply (15). If $\widetilde{\mathcal{S}}^{(k)}(\lambda) \neq \emptyset$, increment k and return to step (2). Otherwise, set $\bar{k}(\lambda) = k$ and retain the terminal variable vector $y_t^\dagger(\lambda)$ and the terminal fitted composite disturbance vector $\widehat{u}_t^\dagger(\lambda) = \widehat{u}_t^{(\bar{k}(\lambda))}$.
- (6) **Compare terminal systems.** Repeat steps (1)–(5) for each value of the model complexity parameter λ in the prespecified grid. Select $\widehat{\lambda}_\alpha$ using the OOS forecast loss criterion in Section V, where α is the posterior probability level used in the OOS comparison.
- (7) **Selected variable vector.** The final vector $y_t^\dagger(\widehat{\lambda}_\alpha)$ contains the core variables and the additional variables selected from the large candidate set for the economic application.

During model construction, $\widehat{u}_t^{(k)}$ denotes the fitted composite disturbance vector for the system at iteration k . For each λ , the terminal fitted composite disturbance vector is

$$\widehat{u}_t^\dagger(\lambda) = \widehat{u}_t^{(\bar{k}(\lambda))}.$$

After OOS selection, we write $y_t^\dagger \equiv y_t^\dagger(\widehat{\lambda}_\alpha)$ and refer to the resulting SVAR as the selected system.

The framework separates three components that are often conflated. The economic question, core variables, and identifying restrictions define the initial economic analysis, whose

information set may be incomplete. For each value of the model complexity parameter, the iterative procedure constructs a terminal information set. The OOS forecast loss criterion then selects among the resulting terminal systems. The economic question, model construction, and OOS selection therefore play distinct roles.

IV. THEORETICAL PROPERTIES

Section III defines a sequence of current index sets, posterior mode estimates, fitted composite disturbances, auxiliary optimization problems, and index set updates. In this section, we establish that this construction is well defined, terminates after finitely many updates, and produces a unique terminal variable vector. We also derive the exact stopping condition and establish invariance to changes in the units of the remaining candidates and stability with respect to small changes in the fitted composite disturbances. Throughout this section, $\bar{k}$ denotes the terminal iteration for a fixed value of the model complexity parameter λ . All results are conditional on the observed data under the prior $\pi(\theta)$.

Assumption 1. The standard deviations $\hat{\sigma}_i$ used to standardize the variables are strictly positive for all $i = 1, \dots, n$. For every index set reached by the procedure, the argmax set in (8) is nonempty, and the auxiliary optimization problem in (11) has a unique minimizer for every $q \in \mathcal{Q}$.

The uniqueness requirement for the auxiliary minimizer concerns the coefficient vector rather than only its fitted values. This distinction is necessary because the update in (14) depends on which coefficients are zero. The nonempty argmax set and the prespecified deterministic rule stated after (8) determine one posterior mode and hence one fitted composite disturbance vector at every iteration. Under Assumption 1, the following two propositions establish convergence and uniqueness.

Proposition 1. *For each fixed λ , the update rule generates*

$$\mathcal{J}^{(0)} \subseteq \mathcal{J}^{(1)} \subseteq \dots \subseteq \mathcal{N}.$$

Once a variable enters the current system, it remains in every subsequent system. The terminal iteration satisfies $\bar{k} \leq n - n^{(0)}$, and the procedure estimates at most $n - n^{(0)} + 1$ current systems.

Proof. See Appendix S1. □

Proposition 2. *Each fixed value of λ determines a unique sequence $\{\mathcal{J}^{(k)}\}_{k=0}^{\bar{k}}$, a unique terminal index set $\mathcal{J}^\dagger(\lambda)$, and a unique terminal variable vector $y_t^\dagger(\lambda)$.*

Proof. See Appendix S1. □

Proposition 2 establishes that the mapping $\lambda \mapsto y_t^\dagger(\lambda)$ is well defined for the OOS comparison in Section V. It does not imply that terminal vectors associated with different model

complexity parameters are nested. Each update changes the fitted composite disturbances used in the next auxiliary optimization, and the construction paths may therefore differ across values of λ .

For the remaining results, let $T_{\mathcal{L}} = T - \mathcal{L}$ and define the stacked fitted composite disturbances and the standardized regressor matrix by

$$\widehat{U}^{(k)} = \begin{bmatrix} (\widehat{u}_{\mathcal{L}+1}^{(k)})' \\ \vdots \\ (\widehat{u}_T^{(k)})' \end{bmatrix}, \quad X^{(k)} = \begin{bmatrix} (x_{\mathcal{L}+1,\text{std}}^{(k)})' \\ \vdots \\ (x_{T,\text{std}}^{(k)})' \end{bmatrix}.$$

Thus, $\widehat{U}^{(k)} \in \mathbb{R}^{T_{\mathcal{L}} \times n^{(k)}}$ and $X^{(k)} \in \mathbb{R}^{T_{\mathcal{L}} \times p_k}$, where $p_k = (\mathcal{L} + 1)\tilde{n}^{(k)}$. Let $\widehat{\mathbf{u}}_q^{(k)}$ denote column q of $\widehat{U}^{(k)}$, let $\mathbf{1}$ be the $T_{\mathcal{L}} \times 1$ vector of ones, and define

$$M = I_{T_{\mathcal{L}}} - \frac{1}{T_{\mathcal{L}}} \mathbf{1}\mathbf{1}'.$$

The matrix M centers the dependent variable and regressors. The intercept is part of the original regression. After centering, the auxiliary optimization can be written without the intercept, with no change in the estimated coefficients on the regressors.

Lemma 1. *For a fixed equation q and coefficient vector β_q , the minimizing intercept in (11) is*

$$d_q(\beta_q) = \frac{1}{T_{\mathcal{L}}} \mathbf{1}' (\widehat{\mathbf{u}}_q^{(k)} - X^{(k)} \beta_q).$$

Substitution of this intercept into (11) yields the equivalent problem

$$\min_{\beta_q} \left\{ \frac{1}{2T_{\mathcal{L}}} \|M\widehat{\mathbf{u}}_q^{(k)} - MX^{(k)}\beta_q\|_2^2 + \lambda \|\beta_q\|_1 \right\}. \quad (16)$$

Proof. See Appendix S1. □

Lemma 1 allows the remaining results to use the centered auxiliary optimization without altering the estimated coefficients that determine which candidates are admitted. By (14), no remaining candidate is admitted at iteration k when all coefficient blocks are zero. The next proposition states an equivalent condition based on the standardized remaining candidates and the fitted composite disturbances.

Proposition 3. *Under Assumption 1, no remaining candidate is admitted at iteration k , that is,*

$$\widetilde{\mathcal{S}}^{(k)}(\lambda) = \emptyset,$$

if and only if

$$\max_{q \in \mathcal{Q}} \left\| \frac{1}{T_{\mathcal{L}}} (X^{(k)})' M \widehat{\mathbf{u}}_q^{(k)} \right\|_{\infty} \leq \lambda. \quad (17)$$

When $\tilde{n}^{(k)} = 0$, the left-hand side of (17) is defined as zero.

Proof. See Appendix S1. □

Proposition 3 gives economic meaning to the zero coefficients returned by the auxiliary optimization. The left-hand side of (17) is the largest absolute average cross-product between any standardized remaining candidate, contemporaneously or at any included lag, and any fitted composite disturbance indexed by $\mathcal{Q}$. When this quantity does not exceed λ , no remaining candidate contains enough information under the auxiliary criterion to expand the current system. At the terminal iteration, (17) holds because termination is defined by $\widetilde{\mathcal{S}}^{(k)}(\lambda) = \emptyset$. The procedure has therefore exhausted the information among the remaining candidates that is detectable by this criterion at the given value of λ . This result connects the optimization stopping rule to the economic purpose of constructing the SVAR information set.

The preceding results establish finite termination, uniqueness, and an exact characterization of stopping. The next two propositions establish that the update rule is invariant to the units of the remaining candidates and stable under sufficiently small changes in the fitted composite disturbances.

Proposition 4. *Under Assumption 1, fix an iteration k , the current system, and its fitted composite disturbances. Replace each remaining candidate $y_{i,t}$ by $\alpha_i + d_i y_{i,t}$, where $d_i \neq 0$, and recompute its standardization moments. The selected set $\widetilde{\mathcal{S}}^{(k)}(\lambda)$ is unchanged.*

Proof. See Appendix S1. □

Proposition 4 ensures that, for a given iteration and fixed fitted composite disturbances, the variables admitted by the procedure do not depend on arbitrary choices of measurement units. The next proposition turns from changes in measurement units to changes in the fitted composite disturbances. It gives a finite sample condition under which sufficiently small changes in these disturbances leave the selected set $\widetilde{\mathcal{S}}^{(k)}(\lambda)$ unchanged for a given iteration.

For equation q , let $\widehat{\beta}_q^{(k)}$ be the unique solution of (16), and define its nonzero coefficient index set by

$$\mathcal{A}_q^{(k)} = \left\{ h : \widehat{\beta}_{q,h}^{(k)} \neq 0 \right\}.$$

Write $\widetilde{X}^{(k)} = MX^{(k)}$, $\widetilde{\mathbf{u}}_q^{(k)} = M\widehat{\mathbf{u}}_q^{(k)}$, and

$$\widehat{\mathbf{v}}_q^{(k)} = \widetilde{\mathbf{u}}_q^{(k)} - \widetilde{X}^{(k)}\widehat{\beta}_q^{(k)}.$$

Assumption 2. For every equation q , the matrix $T_{\mathcal{L}}^{-1}(\widetilde{X}_{\mathcal{A}_q^{(k)}}^{(k)})'\widetilde{X}_{\mathcal{A}_q^{(k)}}^{(k)}$ is positive definite when $\mathcal{A}_q^{(k)}$ is nonempty. There exist constants $b_q > 0$ and $\delta_q > 0$ such that

$$\min_{h \in \mathcal{A}_q^{(k)}} |\widehat{\beta}_{q,h}^{(k)}| \geq b_q$$

when $\mathcal{A}_q^{(k)}$ is nonempty, and

$$\max_{h \notin \mathcal{A}_q^{(k)}} \left| \frac{1}{T_{\mathcal{L}}} (\widetilde{X}_h^{(k)})' \widehat{\mathbf{v}}_q^{(k)} \right| \leq \lambda - \delta_q$$

when the complement of $\mathcal{A}_q^{(k)}$ is nonempty.

The constants b_q and δ_q define the selection margins for equation q . When nonzero estimated coefficients are present, the first margin is a positive lower bound on their absolute values and measures how far the selected coefficients are from zero. When unselected regressors are present, the second margin is a positive lower bound on the gap between λ and the largest absolute average cross-product between an unselected regressor and the residual $\hat{\mathbf{v}}_q^{(k)}$. It measures how far an unselected regressor is from meeting the threshold for a nonzero coefficient. The following proposition and corollary establish stability when both relevant selection margins are positive.

Proposition 5. *Under Assumptions 1 and 2, there exists $\epsilon_k > 0$ such that replacing each fitted composite disturbance vector $\hat{\mathbf{u}}_q^{(k)}$ by $\hat{\mathbf{u}}_q^{(k)} + h_q$ with $\|h_q\|_2 < \epsilon_k$ leaves every nonzero coefficient index set $\mathcal{A}_q^{(k)}$ unchanged. Hence, the selected variable set $\tilde{\mathcal{F}}^{(k)}(\lambda)$ and the next index set $\mathcal{J}^{(k+1)}$ are unchanged.*

Proof. See Appendix S1. □

Corollary 1. *Suppose Assumptions 1 and 2 hold at every iteration of the construction path for a given λ . For each iteration k and equation q , replace $\hat{\mathbf{u}}_q^{(k)}$ by*

$$\hat{\mathbf{u}}_q^{(k)} + h_q^{(k)}, \quad \|h_q^{(k)}\|_2 < \epsilon_k,$$

where ϵ_k is the stability radius in Proposition 5. Then the entire sequence of index sets and the terminal variable vector $y_t^\dagger(\lambda)$ remain unchanged.

Proof. See Appendix S1. □

Under Assumption 1, the posterior mode exists and the auxiliary minimizers are unique. The prespecified deterministic rule stated after (8) selects one posterior mode and hence determines one fitted composite disturbance vector at every iteration. Proposition 1 establishes that the procedure terminates after finitely many iterations, and Proposition 2 establishes that each value of λ determines a unique construction path and terminal variable vector. Proposition 3 gives economic meaning to termination: no remaining candidate contains enough information under the auxiliary criterion to expand the current system. Proposition 4 establishes that the admission decision at a given iteration does not depend on the units of measurement of the remaining candidates.

Under Assumption 2, Proposition 5 and Corollary 1 have practical importance for implementing the algorithm. The fitted composite disturbances are computed from posterior mode estimates, which may vary slightly with numerical tolerances, rounding, or optimization routines. Proposition 5 establishes that such small changes do not alter the variables admitted at a given iteration when the selection margins hold.

Corollary 1 establishes the corresponding result for the full procedure. A small change at an early iteration could otherwise alter the next system, later fitted disturbances, and

the final selected variables. The corollary rules out this propagation when the perturbations remain within the stability radius at each iteration. The complete construction path and the terminal variable vector then remain unchanged. These results establish the local reproducibility and practical reliability of the algorithm, an important property in applied work where numerical tolerances are unavoidable.

V. OOS SELECTION OF THE MODEL

In this section, we use the OOS forecast performance of the core variables to select the model complexity parameter $\hat{\lambda}_\alpha$ from the candidate values of λ . Given the terminal map $\lambda \mapsto y_t^\dagger(\lambda)$ established in Sections III and IV, each value of λ produces a unique terminal variable vector. The model complexity parameter governs admission in the auxiliary optimization; each admission alters the system and the fitted composite disturbances at later iterations. Terminal system size varies with λ . Our model-construction methodology therefore does not favor small or large systems *a priori*.

V.1. Terminal systems and the timing convention. Let Λ be the finite prespecified grid of model complexity parameter values. For each $\lambda \in \Lambda$, Sections III and IV deliver the terminal index set $\mathcal{J}^\dagger(\lambda)$, the terminal variable vector $y_t^\dagger(\lambda)$, and its dimension $n^\dagger(\lambda)$. Proposition 2 ensures that these objects are uniquely determined. The terminal systems generated by different values of λ need not be nested, and their sizes need not be monotone in λ . We compare them by the OOS forecast loss of the core variables.

Every terminal system is constructed from the common estimation sample through T . The validation period does not enter the construction of any terminal index set. Gu et al. (2020) use forecast performance over a validation sample to tune their hyperparameters. We follow the same training-validation logic to select our model complexity parameter $\hat{\lambda}_\alpha$ and hence the terminal system.⁴

Let θ_λ denote the parameters of the terminal system indexed by λ , and let $Y_T^\dagger(\lambda)$ denote its observations through T . The posterior distribution

$$p_\lambda(\theta_\lambda \mid Y_T^\dagger(\lambda))$$

is computed for each terminal system. Let $s = 1, \dots, S$ index posterior draws $\theta_\lambda^{(s,T)}$. For each posterior draw, we construct the OOS forecast path recursively using the data available in the validation period and compute the corresponding OOS forecast loss.

In our paper, “out of sample (OOS)” refers to the forecasts. The posterior distribution and the terminal system remain fixed after T . At each forecast origin, the forecast uses

⁴Unlike an expanding-window exercise that re-estimates the model as new data arrive (Giannone et al., 2021, for example), we obtain the posterior draws from data through T and hold them fixed throughout the validation period. The forecasts are updated recursively as new observations become available. Our purpose is to compare terminal systems, not to mimic a real-time forecaster who re-estimates the model as new data arrive.

the data available at that time. Once the realization is observed, it is used to compute the forecast error and enters the information set for the next forecast. The validation period serves one purpose: the resulting OOS forecast losses select $\hat{\lambda}_\alpha$.

Let $T + 1, \dots, T + R$ be the validation period, let H be the maximum forecast horizon, and let $\mathcal{O} \subseteq \{T, \dots, T + R - H\}$ be a prespecified set of $F = |\mathcal{O}|$ forecast origins. At an origin $\tau \in \mathcal{O}$, forecasts condition on observations through τ , but the posterior distribution remains based on the estimation sample ending at T . The validation observations update the conditioning values in the forecast recursion and provide the realizations used to compute forecast errors. They do not change the posterior distribution. Every terminal system is evaluated at the same forecast origins, horizons, and realizations.

V.2. OOS forecast loss for the core variables. Let $J = n^{(0)}$ be the number of core variables, placed first in y_t , and let $y_{j,t}$ denote core variable j for $j = 1, \dots, J$. For a core variable represented in log levels, the target is its growth rate. For a core variable represented in levels, the target remains in levels. Let $\mathcal{J}_{\log} \subseteq \{1, \dots, J\}$ index the core variables represented in log levels and define

$$z_{j,t} = \begin{cases} y_{j,t} - y_{j,t-1}, & j \in \mathcal{J}_{\log}, \\ y_{j,t}, & j \notin \mathcal{J}_{\log}. \end{cases} \quad (18)$$

The same transformation is applied to realized values and forecasts for every terminal system. Forecast errors are therefore evaluated in prespecified economic units: growth rates for variables represented in log levels and levels for all other variables, including percentage points for rates and spreads.

For posterior draw s , let $y_{j,\tau+h|\tau}^{(s,T)}(\lambda)$ denote the h step ahead forecast of core variable j from the terminal system indexed by λ . Set $y_{j,\tau|\tau}^{(s,T)}(\lambda) = y_{j,\tau}$ and define

$$z_{j,\tau+h|\tau}^{(s,T)}(\lambda) = \begin{cases} y_{j,\tau+h|\tau}^{(s,T)}(\lambda) - y_{j,\tau+h-1|\tau}^{(s,T)}(\lambda), & j \in \mathcal{J}_{\log}, \\ y_{j,\tau+h|\tau}^{(s,T)}(\lambda), & j \notin \mathcal{J}_{\log}. \end{cases}$$

Define the corresponding realization by

$$z_{j,\tau+h} = \begin{cases} y_{j,\tau+h} - y_{j,\tau+h-1}, & j \in \mathcal{J}_{\log}, \\ y_{j,\tau+h}, & j \notin \mathcal{J}_{\log}. \end{cases}$$

The squared forecast error is

$$\text{SE}_{j,h,\tau}^{(s,T)}(\lambda) = \left[z_{j,\tau+h|\tau}^{(s,T)}(\lambda) - z_{j,\tau+h} \right]^2.$$

The absolute forecast error is

$$\text{AE}_{j,h,\tau}^{(s,T)}(\lambda) = \left| z_{j,\tau+h|\tau}^{(s,T)}(\lambda) - z_{j,\tau+h} \right|.$$

The OOS forecast loss for posterior draw s averages the squared forecast errors over core variables, horizons, and forecast origins:

$$\mathcal{L}_{\text{OOS}}^{(s,T)}(\lambda) = \frac{1}{JHF} \sum_{j=1}^J \sum_{h=1}^H \sum_{\tau \in \mathcal{O}} \text{SE}_{j,h,\tau}^{(s,T)}(\lambda). \quad (19)$$

For each λ , the draws $\{\mathcal{L}_{\text{OOS}}^{(s,T)}(\lambda)\}_{s=1}^S$ form the empirical posterior distribution of the OOS forecast loss. We also compute the OOS forecast loss using absolute forecast errors. In that case, we replace $\text{SE}_{j,h,\tau}^{(s,T)}(\lambda)$ with $\text{AE}_{j,h,\tau}^{(s,T)}(\lambda)$ in (19).

V.3. The reference system and posterior OOS forecast loss differences. The comparison begins with a reference terminal system. For each $\lambda \in \Lambda$, define the posterior mean OOS forecast loss by

$$\overline{\mathcal{L}}_{\text{OOS}}^{(T)}(\lambda) = \frac{1}{S} \sum_{s=1}^S \mathcal{L}_{\text{OOS}}^{(s,T)}(\lambda).$$

The reference model complexity parameter satisfies

$$\lambda^* \in \arg \min_{\lambda \in \Lambda} \overline{\mathcal{L}}_{\text{OOS}}^{(T)}(\lambda). \quad (20)$$

A prespecified ordering of Λ resolves ties in (20). The posterior mean of the OOS forecast loss determines only the reference system.⁵ The final choice of λ uses the posterior distribution of OOS forecast loss differences as follows.

For $\lambda \neq \lambda^*$, let $s = 1, \dots, S$ index independent pairs of posterior draws, with one draw from the terminal system indexed by λ and one draw from the reference system indexed by λ^* . The independent pairing defines a prespecified product-posterior comparison convention for models with different parameter spaces. It is not implied by a joint posterior over the models. Define

$$\Delta^{(s,T)}(\lambda) = \mathcal{L}_{\text{OOS}}^{(s,T)}(\lambda) - \mathcal{L}_{\text{OOS}}^{(s,T)}(\lambda^*), \quad s = 1, \dots, S. \quad (21)$$

The paired differences

$$\{\Delta^{(s,T)}(\lambda)\}_{s=1}^S$$

form the empirical posterior distribution of the OOS forecast loss difference. For every $\lambda \in \Lambda$ and for each loss, let $\widehat{F}_\lambda$ be the empirical distribution function of the S posterior OOS forecast loss differences. Define

$$q_{(1-\alpha)/2}(\lambda) = \inf \left\{ x : \widehat{F}_\lambda(x) \geq \frac{1-\alpha}{2} \right\}, \quad q_{(1+\alpha)/2}(\lambda) = \inf \left\{ x : \widehat{F}_\lambda(x) \geq \frac{1+\alpha}{2} \right\}$$

for $\alpha \in (0, 1)$. Write

$$\text{PI}_\alpha[\Delta(\lambda)] = [q_{(1-\alpha)/2}(\lambda), q_{(1+\alpha)/2}(\lambda)]$$

⁵The posterior mean is the Bayesian expected-loss criterion for choosing the reference system, provided that the posterior expectation exists. This criterion applies to OOS forecast losses based on either squared or absolute forecast errors.

for the equal-tailed posterior probability interval of level α . The final selection rule compares terminal systems through the posterior distribution of their OOS forecast loss differences relative to λ^* .

The resulting probability interval is conditional on the terminal construction paths, the loss-specific reference system, the realized validation outcomes, and the independent product coupling. We use it as an algorithmic posterior comparison device, not as a posterior probability over model space.

V.4. Posterior comparison and the selection rule. At probability level α , define the sets of terminal systems that are credibly better than the reference system and credibly indistinguishable from it in forecast performance by

$$\begin{aligned}\Lambda_\alpha^- &= \{\lambda \in \Lambda \setminus \{\lambda^*\} : q_{(1+\alpha)/2}(\lambda) < 0\}, \\ \Lambda_\alpha^0 &= \{\lambda^*\} \cup \{\lambda \in \Lambda \setminus \{\lambda^*\} : q_{(1-\alpha)/2}(\lambda) \leq 0 \leq q_{(1+\alpha)/2}(\lambda)\}.\end{aligned}$$

A system in Λ_α^- is classified as credibly better than the reference system at probability level α . A system in Λ_α^0 is classified as credibly indistinguishable from the reference because its equal-tailed interval contains zero. A system whose interval lies above zero is classified as credibly worse and is not retained.

The comparison set gives priority to terminal systems that are credibly better than the reference system.

$$\Lambda_\alpha = \begin{cases} \Lambda_\alpha^-, & \Lambda_\alpha^- \neq \emptyset, \\ \Lambda_\alpha^0, & \Lambda_\alpha^- = \emptyset. \end{cases} \quad (22)$$

The explicit inclusion of λ^* in Λ_α^0 reflects that the OOS forecast loss difference between the reference system and itself is identically zero. By construction, therefore, Λ_α is nonempty.

The credibly-indistinguishable classification depends on the loss function. Squared error places greater weight on large forecast misses, whereas absolute error is less sensitive to them (Gneiting, 2011). We therefore apply the preceding construction separately under squared and absolute forecast errors. This application yields the comparison sets $\Lambda_\alpha^{\text{SE}}$ and $\Lambda_\alpha^{\text{AE}}$, each relative to its loss-specific reference system. Define the combined comparison set by

$$\Lambda_\alpha^\cup = \Lambda_\alpha^{\text{SE}} \cup \Lambda_\alpha^{\text{AE}}.$$

The two losses provide alternative admissibility criteria. We take their union because our objective is to retain the largest terminal system that satisfies the posterior comparison criterion under either loss.

Define

$$n_{\alpha,\cup}^{\max} = \max_{\lambda \in \Lambda_\alpha^\cup} n^\dagger(\lambda), \quad \Lambda_{\alpha,\cup}^{\max} = \{\lambda \in \Lambda_\alpha^\cup : n^\dagger(\lambda) = n_{\alpha,\cup}^{\max}\}.$$

Let $\prec$ be a prespecified strict total order on the grid Λ . The selected model complexity parameter is

$$\hat{\lambda}_\alpha = \min_{\prec} \Lambda_{\alpha, \cup}^{\max}. \quad (23)$$

If $\Lambda_{\alpha, \cup}^{\max}$ is a singleton, its sole member is selected. If several systems share the largest dimension, the prespecified ordering $\prec$ resolves the tie. Because terminal systems need not be nested, the largest system in this rule is the system with the largest dimension, not a system that contains every other retained system.

We set the probability level to $\alpha = 0.68$ in the empirical analysis, following the reporting convention of [Sims and Zha \(1999\)](#). The 68% intervals provide sharper discrimination than the 90% and 95% intervals.⁶ Our zero-containing 68% comparison set parallels the band-based logic of the one-standard-error rule of cross validation. That rule selects the most parsimonious model whose estimated prediction error lies within one standard error of the minimum ([Breiman et al., 1984](#); [Hastie et al., 2009](#)). Our rule instead gives priority to credible improvement and otherwise retains the largest system credibly indistinguishable from the reference, because our objective is information retention rather than predictive parsimony.

V.5. Existence and uniqueness of the OOS selection. This subsection establishes that the OOS forecast loss criterion returns exactly one model complexity parameter, and that the mapping from the data to the selected system exists and is unique.

Assumption 3. The posterior probability level satisfies $\alpha \in (0, 1)$, and the following conditions hold.

- (i) The grid Λ is finite and nonempty, and is endowed with a strict total order $\prec$.
- (ii) For every $\lambda \in \Lambda$, the terminal variable vector $y_t^\dagger(\lambda)$ is unique, so the terminal dimension $n^\dagger(\lambda) = |\mathcal{J}^\dagger(\lambda)|$ is well defined.
- (iii) The posterior of each terminal system is based on data through T . The posterior draws are maintained over the validation period, and each squared-error or absolute-error OOS forecast loss in (19) has a finite value for every draw $s = 1, \dots, S$ and every $\lambda \in \Lambda$.
- (iv) For every $\lambda \in \Lambda$ and for each loss, the S posterior draws uniquely determine the empirical quantiles $q_{(1-\alpha)/2}(\lambda)$ and $q_{(1+\alpha)/2}(\lambda)$ used as the interval endpoints.

Condition (ii) is part of Proposition 2. Conditions (i), (iii), and (iv) are part of the implemented procedure: a finite grid with a fixed order, a fixed set of posterior draws, finite OOS forecast loss values, and a fixed rule for reading quantiles off those draws.

The OOS selection is constructed one loss at a time. The next lemma establishes that the entire per-loss construction—the reference system, the credibly better set, the credibly

⁶We report the 90% and 95% intervals as diagnostics. At these wider levels, the sign of the loss difference is unresolved for nearly all terminal systems in our applications.

indistinguishable set, and the retained set Λ_α defined in (22)—is well defined, unique, and nonempty. Proposition 6 then applies this construction to the two losses and combines the resulting retained sets.

Lemma 2. *Let $g \in \{\text{SE}, \text{AE}\}$ denote either the squared or the absolute forecast loss. Let λ_g^* , $\Lambda_\alpha^{g,-}$, and $\Lambda_\alpha^{g,0}$ denote the corresponding loss-specific objects, and let Λ_α^g denote the retained set obtained by applying the rule in (22) to these two sets. Under Assumption 3, λ_g^* and the sets $\Lambda_\alpha^{g,-}$, $\Lambda_\alpha^{g,0}$, and Λ_α^g exist and are unique. Moreover, $\lambda_g^* \in \Lambda_\alpha^{g,0}$, and Λ_α^g is nonempty.*

Proof. See Appendix S1. □

Proposition 6. *Under Assumption 3, the loss-specific retained sets $\Lambda_\alpha^{\text{SE}}$ and $\Lambda_\alpha^{\text{AE}}$, the combined retained set $\Lambda_\alpha^\cup$, the maximal dimension $n_{\alpha,\cup}^{\max}$, the maximal set $\Lambda_{\alpha,\cup}^{\max}$, and the selected model complexity parameter $\hat{\lambda}_\alpha = \min_{\prec} \Lambda_{\alpha,\cup}^{\max}$ in (23) exist and are unique. Consequently, the selected terminal system $y_t^\dagger(\hat{\lambda}_\alpha)$ exists and is uniquely determined by the data and the prespecified procedure.*

Proof. See Appendix S1. □

Proposition 6 completes the link between model construction and OOS selection. For each loss, the reference system minimizes the posterior mean OOS forecast loss. The posterior distribution of OOS forecast loss differences then identifies terminal systems that are credibly better than the reference system or credibly distinguishable from it. The two loss-specific retained sets are combined, and the selection rule retains the terminal system with the largest dimension in their union. Together with Propositions 1 and 2, this result ensures that each model complexity parameter produces one terminal system and that the OOS selection returns one model complexity parameter and hence one selected terminal system.

We next apply the methodology to two of the most studied questions in the SVAR literature: the effects of credit shocks and the effects of monetary policy shocks. We demonstrate the methodology under two classic identification methods: recursive identification in the household credit application and proxy identification in the monetary policy application. The reported impulse-response bands are conditional on the system used to estimate each response. They incorporate posterior parameter uncertainty within that system but do not average across alternative information sets or reporting systems.

VI. HOUSEHOLD CREDIT AND HOUSING PRODUCTION

This application asks whether the boom-and-decline response of output to a household credit shock survives the systematic construction of the SVAR information set. MSV find in a panel of 30 countries that output initially rises after a household credit shock and later declines. Their proxy SVAR gives the shock a credit supply interpretation. Using a 10-variable model with monthly U.S. data, BPSS confirm the same boom-and-decline finding but attribute the later decline to endogenous monetary policy tightening.

The two papers therefore agree on the central impulse response but differ in their interpretation of the later decline. BPSS argue that the interpretation of a household credit shock depends on the information set. Their larger system includes interest rates, money, commodity prices, and credit spreads, but its composition is chosen by the researcher. This leaves open the model construction question: which additional variables belong in the SVAR?

We address this question under MSV’s recursive identification. Output, business credit, and household credit form the core system because they are common to the MSV and BPSS analyses. Starting from this core, the procedure constructs terminal systems from a large candidate set, and OOS forecast performance selects the final information set. The selected system adds several measures of housing production and separates a household credit shock from a housing production shock.

VI.1. Data. Our monthly candidate data set contains about 145 series used in this application and in the monetary policy transmission application discussed in Section VII. Most series are obtained directly from Federal Reserve Economic Data (FRED), maintained by the Federal Reserve Bank of St. Louis. The remaining series are obtained from the public website of the Federal Reserve Board. We retain only series available over the full sample required by each application. This requirement keeps the estimation sample fixed as variables enter the model during construction. For both applications, we report the FRED series ID whenever one is available.

The data transformations follow standard practice in macroeconomic VAR analysis and place the variables on scales suitable for estimation and economic interpretation. Variables measured as indexes, counts, flows, or stocks enter in log levels. These variables include incomes, production, employment, price indexes, credit, money, housing permits, and housing starts. Log transformations reduce differences in scale across these series, and their impulse responses can be interpreted approximately as percentage changes. Interest rates, term and credit spreads, unemployment rates, capacity utilization rates, and other variables measured as rates or shares enter in percentage points. Their impulse responses retain those units.

VI.2. The economic question and the core evidence. MSV begin with a 3-variable recursive VAR in log real GDP, nonfinancial firm debt in percent of lagged GDP, and household debt in percent of lagged GDP. The recursively identified household credit shock raises household debt. GDP initially rises and then falls, even though household debt remains above its pre-shock level throughout the ten-year horizon. MSV state that this first exercise describes the dynamic relation and is not intended to establish a causal relationship. They then use a low mortgage spread as an external instrument for a household credit supply shock. The proxy SVAR produces the same boom and decline of output. The central finding therefore does not rest on the external instrument. It is already present under the recursive identification, which assigns the rise and fall of output to a household credit shock.

BPSS examine whether the same finding survives in a richer monthly U.S. system. Their 10-variable SVAR includes industrial production, prices, household and business credit, money, the federal funds rate, commodity prices, a term spread, and two measures of financial stress. They identify the shocks through heteroskedasticity. Their household credit shock produces a small initial rise in industrial production and prices, followed by a later decline in industrial production and an increase in the federal funds rate. Neither measure of financial stress widens persistently. BPSS therefore attribute the later output decline to endogenous monetary tightening rather than to an inevitable reversal of credit supply or to financial stress. This interpretation does not overturn MSV’s credit supply interpretation. It identifies endogenous monetary policy as an additional channel behind the later decline.

Figure 1 replicates the MSV evidence for the United States using different data, a different sample, and monthly rather than annual frequency. We use monthly data from 1960M1 through 2012M12, which match the calendar span of MSV’s annual panel, 60 monthly lags to preserve the five-year lag span of their VAR, and the monthly Sims–Zha prior.⁷ We follow the BPSS variable definitions because our analysis compares the variables selected by our procedure with the variables chosen by BPSS.

Household credit remains above its pre-shock level throughout the forecast horizon. The posterior median of industrial production rises for about two years and then declines. Although the data, sample, frequency, and credit measures differ from those in MSV’s VAR analysis, the same initial output boom and later decline continue to hold. This is the core evidence that survives BPSS’s larger system despite their measured challenge to MSV’s interpretation. The remaining question is whether both information sets omit a real activity margin that moves with household credit and helps generate the initial output boom.

VI.3. Identification and model construction. We maintain the recursive identification used by MSV throughout our model construction and OOS selection. The three core variables are ordered as industrial production, business credit, and household credit, corresponding to MSV’s ordering of real GDP, nonfinancial firm debt, and household debt. The household credit shock is the third shock identified by the recursive ordering. The HHC equation may respond contemporaneously to IP and BC, but the HHC shock cannot affect IP or BC on impact.⁸

⁷MSV measure outstanding household and nonfinancial firm credit using annual data from the Bank for International Settlements, including bank and nonbank loans and debt securities, and scale these stocks by GDP. BPSS use monthly commercial bank loan stocks in log levels: consumer and real estate loans for household credit (HHC) and commercial and industrial loans for business credit (BC). Their HHC series includes commercial real estate loans that cannot be separated over the full sample. Our HHC and BC follow BPSS. The MSV measures provide broader sectoral coverage, while the BPSS measures provide the monthly frequency required for our application. Both measure credit stocks rather than flows of new lending.

⁸For the monthly systems, we use the BPSS sample, 1973M1–2015M6, and their lag length of 10 months. We use updated vintages of the underlying data, truncated to the same sample period. We also use the

All variables admitted beyond the core are ordered after IP, BC, and HHC. The expanded systems therefore preserve MSV’s recursive ordering of the three core variables. Shocks to the added variables cannot affect the core variables on impact, although their lags enter the core equations and alter subsequent dynamics. Conversely, the core shocks can affect the added variables on impact. Adding variables also changes the fitted core equations, innovations, and impulse responses. These changes show the importance of model construction. Model construction evaluates the remaining candidates against the disturbances of output, business credit, and household credit, while OOS selection evaluates forecasts of these same variables.

At each iteration, the procedure in Section III estimates the current system and constructs the fitted composite disturbances for the three core equations. A remaining candidate is admitted when its contemporaneous value or included lags contain predictive information for at least one of these disturbances. Admission does not alter the maintained contemporaneous restrictions: admitted variables remain ordered after the three core variables, while their lags are allowed to enter the core equations and affect subsequent dynamics. The enlarged system is then reestimated because the fitted core disturbances change when the information set expands. For each value of λ , this process continues until no remaining candidate is admitted. The OOS criterion in Section V compares the resulting terminal systems using forecasts of the same three core variables (i.e., $J = n^{(0)} = 3$ in (19)).

VI.4. The selected system. The procedure selects a 13-variable terminal system at $\hat{\lambda}_\alpha = 0.1129$ (Table 1). In addition to IP, BC, and HHC, the system contains four housing variables: housing starts in the South and new private housing permits for the Northeast, the West, and the nation. It contains three labor market indicators: average weekly hours in goods producing industries, average weekly hours in manufacturing, and initial claims for unemployment insurance. It also contains three financial indicators: the one year Treasury rate relative to the federal funds rate, the excess bond premium, and the GZ credit spread.⁹

The selected system includes variables that capture two dimensions emphasized by BPSS. The one year Treasury spread measures the relation between a market rate and the policy rate, while the excess bond premium and the GZ spread measure corporate credit conditions. The selected system differs more sharply from BPSS on the real side. It adds three labor

monthly Sims–Zha prior, which is closely related to the prior used by BPSS. This common setup allows us to assess whether the variables chosen by our methodology overlap with those in BPSS.

⁹The construction path begins at iteration 0 with 3 variables: {IP, BC, HHC}. Iteration 1 adds {HRS-G, HS-S, PER-NE, T1Y-FF, EBP, GZS}, producing a 9-variable system. Iteration 2 adds {HRS-M, PER}, producing an 11-variable system. Iteration 3 adds {UI, PER-W}, producing the selected 13-variable system. See the detailed variable descriptions in Table 1. Computation used MATLAB R2023b under Windows 11 Enterprise on an Intel Core Ultra 7 165U processor at 2.10 GHz with 32 GB of RAM. MATLAB had access to 12 CPU cores, although computation was serial. Model construction across all values of λ took 75 seconds. OOS selection based on 2,000 posterior draws took 1,900 seconds, or about 32 minutes. Increasing the number of posterior draws does not change the results.

market indicators and four housing variables. The housing variables form the largest coherent block among the selected additions and provide the new economic margin. Neither MSV nor BPSS includes housing activity in its SVAR.

No housing block is imposed in model construction. Each housing series competes with all other candidates on the same terms. The procedure chooses housing starts in the South, but not national starts or starts in other regions. Among the housing starts series, southern starts are the one retained by the admission criterion at the selected value of λ . By contrast, the selected permit series cover the nation, Northeast, and West, indicating that housing adjustments differ across regions and over time. The geographic labels carry no separate structural interpretation. Together, these series measure several distinct margins of the housing production cycle and contribute predictive information for the fitted composite disturbances in the three core equations.

The selection of housing starts and permits is consistent with a broader literature on housing and credit allocation. [Leamer \(2007, 2015\)](#) place residential construction and housing quantities at the center of the U.S. business cycle. [Rognlie et al. \(2018\)](#) develop a model in which overbuilding of durable capital such as housing can generate a subsequent demand-driven recession. [Müller and Verner \(2024\)](#) show more broadly that the macroeconomic consequences of credit expansions depend on the sector receiving the credit, as nontradable sectors rely more heavily on real-estate-secured borrowing. This literature supports the economic relevance of the housing production margin discovered by our procedure.

VI.5. Household credit is leverage without production. Figure 2 reports the responses to a household credit shock in the selected system. Household credit rises on impact, and its posterior median remains positive throughout the horizon. The 90% credible band excludes zero for most of the horizon. Output does not follow. Its posterior median rises only trivially on impact, and this increase is negligible relative to its later decline. The household credit expansion therefore occurs without a material expansion in output.

The force of this result lies in the central response, not merely in the credible bands. MSV and BPSS interpret the boom and subsequent decline from the estimated response itself. If the selected system merely widened the bands while leaving the initial rise in place, their economic interpretation would survive with less precision. It does not. The posterior median shows no economically meaningful rise in output. This does more than weaken the evidence for an output boom. It removes the boom from the estimated response. The selected information set therefore changes the economic interpretation, not just its statistical precision. Household credit expands without an output boom.

The housing activity responses explain the change in interpretation. The posterior medians of national permits, western permits, northeastern permits, and southern housing starts are negative. The 90% credible bands for national permits, western permits, and southern housing starts lie below zero for substantial portions of the horizon. The response of

northeastern permits is less precise, but its posterior median declines. The GZ spread rises only briefly before returning toward zero. The household credit shock therefore produces a contraction, rather than an expansion, in housing production and no persistent deterioration in corporate credit conditions.

The increase in household credit is instead accompanied by lower residential construction. The aggregate data do not identify the exact use of the additional credit. It may reflect purchases of existing homes, refinancing, home equity extraction, consumer borrowing, or other balance sheet adjustments. These activities can increase outstanding household credit without financing new residential construction—more leverage without more production. Household credit rises, housing production falls, and output shows no meaningful initial increase.

The later output decline is imprecisely estimated in all three exercises. In our selected system, both the 68% and 90% credible bands for the IP response contain zero. In BPSS, the corresponding bands are somewhat tighter but also contain zero during the later decline. In MSV, the 95% confidence interval contains zero over almost the entire ten-year horizon in which the GDP estimate turns negative. The later decline is therefore a feature of the central responses, not a precisely estimated effect.

This imprecision limits what the impulse responses can establish about the source of the later decline. The deleveraging and weak demand interpretation emphasized by MSV and the endogenous monetary tightening interpretation emphasized by BPSS remain consistent with our evidence. Our interpretation does not require output to decline when housing production falls. A decline in housing production works against an output boom, but other components of activity may offset that effect. We therefore remain agnostic about the sign of the later output response. Our main result concerns the initial response: once the selected information set accounts for housing production, a household credit shock no longer produces an output boom in the point estimate, measured by the posterior median.

VI.6. Output rises with housing production. We refer to the recursive innovation associated with national building permits as the housing production shock. The label summarizes its empirical content and does not assign a primitive cause. National permits rise on impact. Permits in the Northeast and West and housing starts in the South rise with them (Figure 3). These joint responses support interpreting the shock as a broad expansion in prospective and current residential construction.

Industrial production rises after the housing production shock. Its 90% credible band remains above zero for about two years, and its 68% band remains above zero longer. Household credit also rises persistently, with both bands above zero over most of the horizon. The posterior median of business credit rises as well, although its response is less precise. Permits signal construction to come, starts record construction under way, and industrial production captures aggregate real activity. Housing activity and output rise together after

the shock and later decline together. This comovement is consistent with the emphasis of [Leamer \(2007, 2015\)](#) on housing activity in the U.S. business cycle.

The contrast with the household credit shock is sharp. The posterior median of industrial production rises substantially after the housing production shock but only trivially after the household credit shock. Household credit rises after both shocks and therefore does not by itself distinguish their output responses. Housing production rises after the housing production shock and falls after the household credit shock. Within the selected system, output rises only in the first case. The comparison shows that the output response changes with the information set; it does not isolate the contribution of any single added variable.

VI.7. Three interpretations. The response to a household credit shock admits of three complementary interpretations. MSV interpret the initial rise in output as a credit supply driven consumption boom and associate the later decline with debt overhang, deleveraging, and macroeconomic frictions. BPSS leave the source of the initial rise largely open, noting that credit and output generally move together, and attribute the later decline to endogenous monetary tightening. Our selected system adds a third interpretation: housing production is an omitted real activity margin that can explain the initial boom in systems that exclude housing starts and permits. These three interpretations can coexist.

MSV and BPSS differ mainly in their explanations of the later output decline. In MSV, the 95% confidence band contains zero over nearly all of the horizon on which the estimated GDP response is negative. In BPSS, both credible bands contain zero during the later decline. Yet both papers develop economic interpretations of that decline. Their interpretations therefore rest on the central impulse responses. In their argument, wide bands reduce precision but do not alter the economic interpretation. In our selected system, the posterior median of output also declines later. This response remains compatible with the debt overhang and deleveraging interpretation of MSV, the endogenous monetary tightening interpretation of BPSS, or both. Our selected system does not distinguish between them, and our main result does not require the later decline to be statistically different from zero.

The same focus on central responses also shapes their interpretation of the initial output rise. MSV interpret its positive central response as a consumption boom driven by credit supply. BPSS also obtain a positive initial response and relate it to the general positive comovement between credit and output, but they do not identify the spending or production margin behind the initial rise. Neither information set includes housing starts or permits. These systems therefore cannot determine whether the initial output rise comes from household credit itself or from residential construction that moves with credit.

Our selected system separates these margins. At the point estimate, measured by the posterior median, the initial response of industrial production to the household credit shock is near zero and economically negligible. This is not a case in which wider bands surround

an output boom. The boom is absent from the estimated response itself. Housing production also falls after the household credit shock. This finding does not imply that falling housing production must cause aggregate output to decline, nor does it rule out support for consumption. It shows that any such support is insufficient to produce a meaningful aggregate output boom when residential construction contracts. Our contribution therefore concerns the source of the initial boom rather than the source of the later decline. Models of household credit and the business cycle should allow residential construction to move separately from household credit. In smaller systems that do not separate these margins, the identified household credit shock may combine movements in credit with movements in housing production.

VI.8. Robustness. The next two sets of exercises establish two distinct results. First, the initial output boom in the original researcher-chosen systems is unaffected by differences in data frequency, variable definitions, updated vintages, and the Sims–Zha prior. Second, the selected 13-variable system produces no initial output boom under the heteroskedasticity identification of BPSS.

First, Figure 1 replicates MSV’s VAR evidence using data that differ from theirs in both frequency and variable definition, as discussed in Section VI.2. We use monthly rather than annual data and the monthly Sims–Zha prior, whereas MSV do not use a Bayesian prior. Figure 4 replicates the IP, BC, and HHC responses to the household credit shock reported by BPSS. We use their 10-variable system and sample period, but updated vintages of the monthly data. We also use the Sims–Zha prior, which is close to but not identical to the BPSS prior.¹⁰ In both cases, the posterior median of output rises initially and later declines. Different data, updated vintages, and the Sims–Zha prior do not remove the initial output boom. The point estimate of the initial rise remains sizable.

Second, the selected system yields the same separation under BPSS’s identification through shock heteroskedasticity. We apply their identifying method to our 13-variable system. Figure 5 shows that a household credit shock raises household credit without producing an output boom, while all four housing activity variables decline.¹¹ The GZ spread remains essentially unchanged. Because this spread measures corporate credit conditions, its muted response indicates that the household credit expansion is not accompanied by a broader deterioration in corporate credit markets.

Figure 6 shows that a housing production shock raises housing permits, housing starts, industrial production, household credit, and business credit. The same separation therefore obtains under BPSS’s identification through shock heteroskedasticity and is not an artifact

¹⁰The full panel of impulse responses from this 10-variable BPSS system is reported in Appendix S3.

¹¹These results hold when we augment the selected system with six BPSS variables absent from that system: the price level, M1, the federal funds rate, the 10Y–3M Treasury spread, the TED spread, and the commodity price index. In particular, there is no initial output boom after the household credit shock.

of the recursive identification. Across both identifying methods, household credit rises after both shocks, but output rises only when housing production expands. This result supports the housing production channel.

VII. MONETARY POLICY TRANSMISSION

This application asks how systematic construction of the SVAR information set changes the estimated transmission of monetary policy under proxy identification. We apply the methodology of Sections III–V to GK’s proxy SVAR. The application further develops a necessary and sufficient identification result for multiple instruments identifying one shock and an anchor-free joint Bayesian method. We first describe GK’s two-stage procedure in our notation and then show how the proxy restrictions identify the monetary policy shock within the general SVAR framework.

The joint approach is not merely an extension or a broader formulation. It is needed for coherent and stable economic results. Both GK’s procedure and ours identify the same monetary policy shock and the same impulse responses. They differ in how they recover these responses from the data. GK’s procedure singles out the residual of one variable as an anchor and measures every response relative to it. Our approach estimates the responses jointly and eliminates the need for such a choice. With a single instrument, the anchor does not matter, and GK use one instrument. With several instruments, however, the anchor generally matters in finite samples. The estimated responses and their credible bands can move with the variable chosen as the anchor, and no economic argument selects one variable over another.

Moving from GK’s single instrument to the multiple instruments of Swanson (2021) makes the joint method necessary for avoiding arbitrary finite-sample dependence on the choice of anchor. The empirical subsection uses the six variables from GK’s Figure 2 as the core variables and the three external instruments of Swanson (2021) to show how the joint approach operates within our system-selection methodology and what it implies for the transmission of monetary policy.

VII.1. Proxy identification within one SVAR system. To avoid burdensome system-specific notation, throughout this section we reuse n , A_0 , A_ℓ , ε_t , Σ , and related symbols for the SVAR system under consideration. With this slight abuse of notation, the system may represent the original core-variable system, a current system generated during model construction, or the selected terminal system, depending on the context. The dimensions and values of these objects are understood to conform to the system under consideration. The proxy identification argument below applies to the system being analyzed.

For every system under consideration, the monetary policy equation is placed first, and ε_{1t} is the monetary policy shock. Let v_t denote the reduced-form innovation of that system,

where the symbol v is used because u is reserved in Section III for the composite disturbance:

$$v'_t = \varepsilon'_t A_0^{-1}, \quad \Sigma \equiv \mathbb{E}(v_t v'_t) = A_0^{-1'} A_0^{-1}. \quad (24)$$

Let e_1 be the first column of I_n and define

$$a_1 \equiv A_0 e_1, \quad s \equiv A_0^{-1'} e_1, \quad s' = e'_1 A_0^{-1} = a'_1 \Sigma. \quad (25)$$

The vector a_1 contains the coefficients in the monetary policy equation. The row vector s' gives the impact responses to the monetary policy shock.

Let m_t be an $n_m \times 1$ vector of centered external instruments, where $n_m \geq 1$. Following the literature, we maintain the following proxy restrictions.

Assumption 4. The instrument vector satisfies

$$\mathbb{E}(m_t \varepsilon_{1t}) = \gamma_m \neq 0_{n_m \times 1}, \quad \mathbb{E}(m_t \varepsilon_{jt}) = 0_{n_m \times 1}, \quad j = 2, \dots, n. \quad (26)$$

Let $\Omega_m \equiv \mathbb{E}(m_t m'_t)$. The matrix Ω_m is positive definite.

Assumption 4 requires the instruments to have predictive content for the policy shock and to be orthogonal to every other structural shock. Define the reduced-form proxy covariance matrix

$$G \equiv \mathbb{E}(v_t m'_t) \in \mathbb{R}^{n \times n_m}. \quad (27)$$

For $i = 1, \dots, n$, let $g_i \in \mathbb{R}^{n_m}$ denote the transpose of the i th row of G , that is, $g_i = \mathbb{E}(m_t v_{i,t})$, and let $W \equiv \Omega_m^{-1}$.

The GK two-stage procedure. GK first estimate the reduced-form VAR and obtain its residuals. They then choose one reduced-form residual as the policy indicator used in the first stage. Let $p \in \{1, \dots, n\}$ index this anchor residual. An anchor p is defined as admissible if and only if $g_p \neq 0_{n_m \times 1}$. Assuming p is admissible, the two-stage least-squares coefficient from regressing $v_{i,t}$ on $v_{p,t}$, using m_t as the instrument vector, is

$$\beta_{i|p} = \frac{g'_p W g_i}{g'_p W g_p}. \quad (28)$$

Following the external-instrument argument of GK, Theorem 1 below expresses their identifying conditions in matrix form. Assumption 4 implies $G = s \gamma'_m$, hence $g_i = \gamma_m s_i$ and

$$\beta_{i|p} = \frac{s_i}{s_p}$$

whenever $s_p \neq 0$. Collect these coefficients in the anchor-specific ratio vector

$$d_p = (\beta_{1|p}, \dots, \beta_{n|p})',$$

which satisfies $\beta_{p|p} = 1$ and $d_p = s/s_p$. The two-stage procedure recovers this ratio representation and then uses the unit-variance normalization of the policy shock to recover its scale:

$$s = \pm \frac{d_p}{\sqrt{d'_p \Sigma^{-1} d_p}}. \quad (29)$$

The sign is chosen by an economic normalization, such as requiring a contractionary policy shock to raise the interest rate on impact.

GK choose the reduced-form residual for the one-year government bond rate as the admissible anchor, estimate the impact ratios relative to that residual, and then recover the scale of the impact vector. They do not examine how the ratio representation changes when a different residual is used as the admissible anchor. The next proposition shows that different anchors change this intermediate representation but recover the same impact vector and the same impulse responses after normalization.

Proposition 7. *Under Assumption 4, every admissible anchor p yields the ratio vector*

$$d_p = \frac{s}{s_p}.$$

For any two admissible anchors p and r ,

$$d_r = \frac{s_p}{s_r} d_p.$$

Thus, different anchors yield different ratio vectors that are proportional to one another. After the scale and sign normalization in (29), every admissible anchor recovers the same impact vector s and the same impulse responses at every horizon.

Proof. See Appendix S2. □

Proposition 7 has a sharp finite-sample implication. With one instrument, the estimate $\hat{G}$ has one column and therefore has rank one whenever it is nonzero. The sample analogue of the proposition then holds exactly in the data: different admissible anchors yield proportional ratio vectors, the same normalized impact vector, and the same impulse responses. With several instruments, Assumption 4 still makes G rank one in population, but $\hat{G}$ need not be rank one and is generally of higher rank in finite samples. Different anchors can then yield ratio vectors that are not proportional, causing the estimated impulse responses and their error bands to depend on the chosen anchor. The joint method introduced next estimates the entire system together. Its likelihood and posterior do not depend on any choice of anchor and therefore avoid this arbitrary finite-sample dependence.

The joint proxy SVAR. Our method estimates the SVAR and the proxy relation jointly under one posterior distribution. The contemporaneous second moments of the joint system satisfy

$$\mathbb{E} \left(\begin{bmatrix} \varepsilon_t \\ m_t \end{bmatrix} \begin{bmatrix} \varepsilon'_t & m'_t \end{bmatrix} \right) = \begin{bmatrix} I_n & e_1 \gamma'_m \\ \gamma_m e'_1 & \Omega_m \end{bmatrix}, \quad (30)$$

where Ω_m is positive definite. The block structure imposes the proxy moments in (26) and preserves the unit-variance normalization in (1). Under joint normality, we represent (30) as an augmented structural system for $\tilde{y}_t = [y'_t, m'_t]'$. The proxy exogeneity conditions become linear, non-cross-equation block restrictions on the augmented coefficient matrices, and the

augmented shock vector is normally distributed with mean zero and identity covariance matrix. We place the Sims–Zha prior on the augmented system and draw from its posterior using the efficient Gibbs sampler of [Waggoner and Zha \(2003\)](#), which applies directly and without modification. Because each Gibbs draw has an invertible augmented contemporaneous matrix, the implied covariance matrix in (30) is positive definite, and the Schur complement condition (32) below holds automatically.

[Caldara and Herbst \(2019\)](#) develop a published model with one proxy for one shock. Their online appendix considers two proxies for one shock but assumes that the two measurement errors are independent. In our notation, this assumption requires $\Omega_m - \gamma_m \gamma_m'$ to be diagonal. It therefore excludes residual covariance between the proxies and requires all covariance between them to arise through their common loading on the shock. Identification does not require this restriction. Applied to our three instruments, it would impose three zero restrictions on the off-diagonal elements of the measurement error covariance matrix. When written as an augmented structural system, their measurement equation also couples the proxy equation to the policy equation through cross-equation proportionality restrictions.

These restrictions fall outside the non-cross-equation class studied by [Waggoner and Zha \(2003\)](#). The posterior simulator of [Caldara and Herbst \(2019\)](#) is therefore a Metropolis-within-Gibbs algorithm with tuned proposals. When the proxy is highly informative, the proposal based on the VAR-only posterior can differ sharply from the joint posterior, causing low acceptance rates. They consequently use a sequential Monte Carlo sampler for their reported applications, which they describe as stable but computationally demanding. Our specification instead leaves Ω_m unrestricted within the class of positive definite matrices and imposes only proxy exogeneity and the Schur complement condition in (32) below; identification further requires $\gamma_m \neq 0$, as in GK and the earlier external-instrument literature. This unrestricted covariance specification nests both the published scalar model of [Caldara and Herbst \(2019\)](#) and their independent-error multiple-proxy extension. Moreover, Proposition 8 below establishes the population equivalence between our joint Bayesian approach and the GK two-stage procedure.

To interpret the remaining restriction, consider the linear projection

$$\varepsilon_{1t} = m_t' \Omega_m^{-1} \gamma_m + w_{1t}, \quad \mathbb{E}(m_t w_{1t}) = 0_{n_m \times 1}. \quad (31)$$

Since $\text{Var}(\varepsilon_{1t}) = 1$, the fraction of the policy shock variance explained by the instruments is

$$\gamma_m' \Omega_m^{-1} \gamma_m,$$

and the variance of the projection residual is

$$\text{Var}(w_{1t}) = 1 - \gamma_m' \Omega_m^{-1} \gamma_m.$$

The covariance matrix in (30) is positive definite when

$$\gamma_m' \Omega_m^{-1} \gamma_m < 1. \quad (32)$$

The projection residual has positive variance if and only if the Schur complement inequality in (32) holds. Together with $\gamma_m \neq 0$, this inequality implies that the instruments provide information about the policy shock without perfectly predicting it. Given that Ω_m is positive definite, it also ensures that the joint covariance matrix in (30) is positive definite.

Utilizing (31) in the first equation of (1) gives

$$y'_t a_1 = c_1 + \sum_{\ell=1}^{\mathcal{L}} y'_{t-\ell} a_{\ell,1} + m'_t \delta_m + w_{1t}, \quad a_{\ell,1} = A_{\ell} e_1, \quad (33)$$

where $\delta_m = \Omega_m^{-1} \gamma_m$. The instruments enter the first equation through the linear projection of the policy shock on m_t . For $j = 2, \dots, n$, Assumption 4 implies $\mathbb{E}(m_t \varepsilon_{jt}) = 0_{n_m \times 1}$. The linear projection of each remaining shock on the instruments is therefore zero, and no instrument term enters the remaining equations.

The external instrument vector m_t joins the SVAR system under consideration through the first equation. By construction, $\mathbb{E}(m_t w_{1t}) = 0_{n_m \times 1}$. For $j = 2, \dots, n$,

$$\mathbb{E}(w_{1t} \varepsilon_{jt}) = \mathbb{E}(\varepsilon_{1t} \varepsilon_{jt}) - \delta'_m \mathbb{E}(m_t \varepsilon_{jt}) = 0,$$

because the model shocks are contemporaneously uncorrelated and Assumption 4 makes the instruments uncorrelated with every remaining shock. Thus, w_{1t} is contemporaneously uncorrelated with the shocks in the remaining equations.

This augmented specification is what we have referred to as the joint system. The coefficient δ_m is estimated together with all other parameters in that system. Theorem 1 below establishes that the joint system identifies all coefficients in the first equation of the SVAR system under consideration and therefore identifies the impulse responses to the first shock at every horizon.¹² No anchor is chosen, and no ratio calculation is needed.

Theorem 1. *Assume that $n \geq 2$, the SVAR system under consideration has an invertible A_0 , and (24) holds. Let*

$$\gamma_m \equiv \mathbb{E}(m_t \varepsilon_{1t}), \quad \Omega_m \equiv \mathbb{E}(m_t m'_t),$$

where Ω_m is positive definite, and maintain the exclusion restrictions

$$\mathbb{E}(m_t \varepsilon_{jt}) = 0_{n_m \times 1}, \quad j = 2, \dots, n.$$

In the absence of additional identifying restrictions on the SVAR system, the joint system identifies all coefficients in the first equation up to their common sign if and only if

$$\gamma_m \neq 0_{n_m \times 1}.$$

¹²Following Theorem 4 of Zha (1999), Proposition S1 in Appendix S2 shows that an orthogonal transformation of columns 2, $\dots$, n leaves the joint covariance restriction, the first equation, and the impulse responses to the first shock unchanged. The identification result is related to the partial-identification framework of Rubio-Ramírez et al. (2010), whose Theorem 2 gives a general sufficient rank condition for identifying one equation. Here the proxy moments yield a necessary and sufficient condition for identifying the first equation and provide explicit formulas for its coefficients and impulse responses.

Equivalently, identification holds if and only if $G \neq 0_{n \times n_m}$, or, equivalently,

$$\text{rank}(G) = \text{rank}(\Sigma^{-1}G) = 1.$$

When these equivalent conditions hold,

$$G = s\gamma'_m, \quad \Sigma^{-1}G = a_1\gamma'_m. \quad (34)$$

For any $q \in \mathbb{R}^{n_m}$ such that $Gq \neq 0_{n \times 1}$,

$$a_1 = \pm \frac{\Sigma^{-1}Gq}{\sqrt{(\Sigma^{-1}Gq)' \Sigma (\Sigma^{-1}Gq)}}, \quad s = \Sigma a_1 = \pm \frac{Gq}{\sqrt{(Gq)' \Sigma^{-1}(Gq)}}. \quad (35)$$

Thus, the joint system identifies c_1 , a_1 , $a_{\ell,1}$ for $\ell = 1, \dots, \mathcal{L}$, and δ_m , up to the common sign normalization of the first equation. It therefore identifies the impulse responses to the first shock at every horizon. The remaining equations are not identified without additional restrictions and are not needed for these impulse responses.

Proof. See Appendix S2. □

The identification result in Theorem 1 applies beyond a single shock. The label “first” is merely a notational convention; any shock can be placed first. More important, identification need not proceed one equation at a time. Multiple shocks can be identified simultaneously. Let $\mathcal{J} \subseteq \{1, \dots, n\}$ index a set of shocks, and for each $j \in \mathcal{J}$ let $m_t^{(j)}$ be a vector of instruments satisfying

$$\mathbb{E}(m_t^{(j)} \varepsilon_{jt}) = \gamma_j \neq 0, \quad \mathbb{E}(m_t^{(j)} \varepsilon_{kt}) = 0, \quad k \neq j.$$

Let m_t stack the instrument vectors $m_t^{(j)}$, let $\Omega = \mathbb{E}(m_t m_t')$, and let $C = \mathbb{E}(\varepsilon_t m_t')$. For population identification, assume $\Omega \succ 0$ together with the relevance and exclusion restrictions above. For estimation under one regular joint Gaussian posterior, assume that

$$\begin{bmatrix} I_n & C \\ C' & \Omega \end{bmatrix} \succ 0,$$

or equivalently $I_n - C\Omega^{-1}C' \succ 0$ and $\Omega - C'C \succ 0$. Under these conditions, Theorem 1 applies to each instrument vector and simultaneously identifies the corresponding shock and equation, the impact vector $s_j = A_0^{-1}e_j$, and the corresponding impulse responses at every horizon, up to a separate sign normalization for each shock. The joint method makes this extension straightforward: the instrument vectors for different shocks are stacked, the corresponding restrictions are imposed on the joint covariance block, and estimation proceeds under one posterior distribution as in a standard SVAR with several identified shocks and the corresponding equations.

We now return to the notation of Theorem 1, where the shock of interest is labeled “first.” The vector q is only an algebraic device for displaying the identified direction. Under Assumption 4, Gq is proportional to s for every q such that $Gq \neq 0_{n \times 1}$. The joint posterior is defined using the full instrument vector and does not depend on any choice of q . The

identified impact vector and the reduced-form dynamics determine the impulse responses to the shock of interest.

Define the reduced-form coefficient matrices

$$B_\ell \equiv A_\ell A_0^{-1}, \quad \ell = 1, \dots, \mathcal{L}, \quad (36)$$

and define the moving-average matrices by

$$\Psi_0 = I_n, \quad \Psi_h = \sum_{\ell=1}^{\min(\mathcal{L}, h)} \Psi_{h-\ell} B_\ell, \quad h \geq 1. \quad (37)$$

The horizon- h impulse-response row to the first shock is

$$\text{IRF}_1(h) = s' \Psi_h = a_1' \Sigma \Psi_h. \quad (38)$$

Theorem 1 therefore identifies $\text{IRF}_1(h)$ up to the common sign normalization for every $h \geq 0$.

Proposition 8. *Under Assumption 4, the GK two-stage procedure based on any admissible anchor and the joint proxy SVAR identify the same impact vector s after the common scale and sign normalization. They therefore identify the same first equation and the same impulse responses at every horizon.*

Proof. See Appendix S2. □

Proposition 8 confirms that the joint method preserves the population object identified by GK. The methods differ in finite-sample estimation. The joint method requires no anchor, estimates the SVAR and the proxy relation under one posterior distribution, and carries uncertainty about both components jointly. It also avoids division by an anchor loading. With one instrument, the normalized GK procedure is anchor invariant even in finite samples. With multiple instruments, the sample covariance matrix generally need not have rank one, and the GK estimates can depend on the residual chosen as the anchor. The joint method remains anchor free. Proposition S2 in Appendix S2 shows that, under its stated regularity conditions, the joint estimator attains the parametric efficiency bound and weakly dominates any regular two-stage estimator of the same object under the same model.

VII.2. Policy surprises and financial conditions. The preceding subsection establishes that the joint proxy SVAR identifies the monetary policy shock, its equation, and its impulse responses without choosing an anchor. We now combine this anchor-free identification with model construction from the large data set described in Section VI to revisit GK's central question: how monetary policy surprises pass through financial markets to prices and real activity. GK use the 6-variable specification in their Figure 2 as the baseline: the one-year Treasury rate, CPI, industrial production, the excess bond premium, the mortgage spread, and the commercial paper spread (see Table 2 for data details). The one-year Treasury rate summarizes both the current policy stance and the expected path of short rates. CPI and industrial production measure the aggregate price and real activity responses. The

three spreads trace distinct credit-cost margins in corporate bond, mortgage, and short-term business credit markets.

These 6 variables define the economic question and therefore serve as the core variables throughout model construction and OOS selection. Our methodology then determines which additional variables enlarge the information set. This approach preserves direct comparability with GK while replacing their researcher-chosen extensions with our model-construction procedure and OOS forecast loss criterion. Identification and information selection perform distinct tasks: the three high-frequency instrument series constructed by [Swanson \(2021\)](#) are used jointly as one instrument vector to identify the monetary policy shock, while the model-construction procedure and OOS forecast loss criterion determine which variables are included to trace its transmission through financial markets to prices and real activity. The model-construction sample runs from 1988:02 through 2012:06. The end date matches GK’s sample endpoint, while the start date is set by the earliest common availability of the three instrument series. The OOS validation sample covers the next 10 years, from 2012:07 through 2022:06.

We treat the three series of [Swanson \(2021\)](#) as one instrument vector for a single monetary policy shock under Assumption 4. Their federal funds rate, forward-guidance, and large-scale-asset-purchase labels describe distinct dimensions of announcement surprises, not separate shocks in our model.

VII.2.1. The economic question and the core evidence. At the center of GK is a fundamental question about monetary policy transmission. In a frictionless benchmark, monetary policy affects private borrowing rates only by altering current and expected future short rates. Term premia and credit spreads do not respond. A modest movement in the path of safe rates should therefore produce a similarly modest movement in private credit costs. With financial frictions, however, a monetary tightening can widen credit spreads. The resulting increase in private borrowing costs can lead to declines in economic activity. The 6-variable core system that GK use as their baseline directly addresses the central question of their paper: “how credit costs respond to exogenous surprises in monetary policy.”

Figure 7 provides our empirical estimates for this core system. The excess bond premium and commercial paper spread rise sharply after the monetary policy shock, with credible bands above zero at both the 68% and 90% levels, and then decline toward zero. The mortgage spread also rises, although its credible bands lie above zero only at the 68% level and in the short run. By contrast, the one-year Treasury rate rises only slightly on impact, and its credible bands include zero at both the 68% and 90% levels. Private credit conditions therefore tighten more clearly than the policy indicator itself. The response of the economy cannot be summarized by a large and precisely estimated movement in the one-year rate. CPI falls persistently, with credible bands below zero at the 90% level for almost two years.

The median response of industrial production also turns negative, but its credible bands lie below zero only at the 68% level, between one and two years after the shock.

The broad economic response supports GK’s central mechanism. A monetary policy shock raises credit spreads across corporate bond, mortgage, and short-term business credit markets while prices and output decline. The rise in borrowing spreads represents a transmission margin beyond the safe rate. It increases the cost of external finance faced by firms and households and puts downward pressures on prices and output. The evidence is particularly strong for the excess bond premium and the commercial paper spread, which identify transmission through both long-term corporate credit and short-term business finance.

At the same time, our results differ from GK in informative respects. In their Figure 2, the one-year rate rises by roughly 20 basis points on impact, industrial production falls significantly after several months, CPI declines but not significantly at the 95% level, and all three credit spreads rise significantly. In our Figure 7, the decline in CPI is more precisely estimated than the decline in industrial production, and the response of the one-year rate is not credibly different from zero. The main agreement between our results and GK’s is the increase in credit spreads across several financial markets. The precise response of the policy indicator and the relative strength of the price and output responses are less robust across the two empirical specifications.

This difference reflects design differences in several respects: the samples differ, the instruments differ, the estimation methods differ, and the error bands differ. GK identify the shock by instrumenting the reduced-form innovation in the one-year Treasury rate with FF4, the high-frequency surprise in the three-month-ahead federal funds futures rate. The one-year rate is the anchor of their construction. We instead use the three high-frequency series of Swanson (2021) jointly as one instrument vector and estimate the proxy relation and the SVAR under one posterior distribution without any anchor. The three series contain policy news associated with the federal funds rate, forward guidance, and large-scale asset purchases.

Our credible bands are wider than those reported by GK. The sample, instruments, model, prior, estimation method, and interval construction all differ, so this cross-study comparison does not identify the source of the difference and is not an efficiency comparison. Within our model, posterior uncertainty about the common proxy direction, the instrument loadings, and the SVAR parameters enters every impulse response. The wider credible bands do not contradict the asymptotic efficiency result in Proposition S2 of Appendix S2, which shows that the joint estimator attains the efficiency bound and weakly dominates regular two-stage estimators under the same model.

Yet, despite these differences in design and statistical uncertainty, the central transmission mechanism through credit markets holds. Our design reproduces the central GK finding: a monetary policy surprise widens credit spreads. The result holds under a different identification, with three instruments jointly and no privileged reduced-form innovation, and without

a large or precise response of the one-year rate. Even without such a response, the excess bond premium and commercial paper spread rise credibly, the mortgage spread is credibly positive at the 68% level, and prices and output are under downward pressures from these tightening credit markets. GK stress that modest movements in the one-year safe rate can produce large movements in credit costs. In that sense, our results strengthen their message: even without a credible movement in the one-year safe rate, contractionary monetary policy still tightens credit markets. Thus, our results establish the crucial fact behind the GK interpretation: monetary policy surprises tighten private credit conditions in ways that are not summarized by the response of the one-year safe rate. This fact highlights an important financial amplification margin, as emphasized by GK.

VII.2.2. Selection. Our model-construction procedure concerns only the monetary policy shock, because that is the single shock identified. It therefore targets one component of the composite disturbance, the policy-equation component $\hat{u}_{1,t}^{(k)}$, which is scalar. As discussed in Sections III and IV, this component changes as variables are admitted. Each admitted variable enlarges the current system. The enlarged system is then reestimated by the joint method of Section VII.1, and $\hat{u}_{1,t}^{(k)}$ changes accordingly. The OOS forecast loss criterion selects a 19-variable system at $\hat{\lambda}_\alpha = 0.0881$: the 6 GK core variables and the 13 additional variables reported in Table 2.¹³

Our methodology selects neither a ladder of two-year, five-year, and ten-year Treasury rates nor several closely related Treasury-minus-federal-funds spreads. This is intuitive because these highly correlated variables add little distinct information for predicting the innovation in the policy equation. The fact that the 19-variable selected system is not nested in GK’s sequence of interest rate extensions is economically informative. The procedure does not merely choose which maturity to append to their baseline. It instead adds variables that measure distinct margins of labor adjustment, housing activity, asset valuation, volatility, credit pricing, liquidity conditions, exchange rates, and commodity prices, none of which can be recovered by moving along a Treasury yield ladder. The selected system therefore expands the empirical content of monetary policy transmission rather than only its interest rate dimension.

The selected variables pass two distinct procedures. They first enter because their contemporaneous values or lags have predictive content for the one fitted composite disturbance of the policy equation in the current system. The resulting terminal system is then compared with the other terminal systems through the OOS forecast loss of the core variables.¹⁴ This

¹³The 19-variable system at $\hat{\lambda}_\alpha = 0.0881$, selected by the OOS forecast loss criterion, reaches its terminal system after one update. The procedure begins with the 6 GK core variables at iteration 0 and adds all 13 additional variables in the first and only update. Across the candidate values of λ , most terminal systems require two updates, some require one, and a few require more than two.

¹⁴Computations were performed in MATLAB R2023b under Windows 11 Enterprise on a computer with an Intel Core Ultra 7 165U processor at 2.10 GHz and 32 GB of RAM. MATLAB had access to 12 CPU

separation matters. The variables are not chosen because they produce preferred impulse responses or because their economic labels fit a prior story. They enter through model construction, and the corresponding terminal system is evaluated by the common OOS criterion before the impulse responses are interpreted.

Model construction and impulse-response reporting serve different purposes. The 13 additional variables enter jointly in the optimization procedure, and their conditional predictive content for the fitted composite disturbance determines their admission. At the terminal iteration, they form the selected 19-variable information set in which the policy equation and monetary policy shock are estimated.

Identification is system specific. Given any core-plus-one reporting system, Theorem 1 identifies the policy shock and its equation for that system under the maintained relevance and exclusion restrictions; the 19-variable dimension is not an identification condition for that given system. The selected information set determines which transmission margins are reported. We therefore follow GK and estimate each core-plus-one system separately because the selected 19-variable system produces credible bands too wide to distinguish the individual transmission margins clearly. Each reported response is conditional on the corresponding reporting system, not on the selected 19-variable system.

VII.2.3. Transmission margins beyond the GK core. The 13 additional variables expand the GK core along six economically distinct margins: labor market adjustment, housing activity, external and commodity conditions, equity valuation and volatility, credit and liquidity conditions, and expected default risk. We take them in turn. Each figure adds one variable from the selected information set to the six-variable core, holds the core fixed, and reports posterior medians with 68% and 90% credible bands.¹⁵ We move from the real-activity and external margins, where the responses show downward pressures with limited precision, to the financial margins, where the responses are credible at the 90% level over the early horizons.

Labor market adjustment. The selected system adds short-duration unemployment, hours in goods-producing industries, and hours in manufacturing (Figure 8). Hours are the intensive margin of labor demand, which firms can vary without changing employment or capacity. Short-duration unemployment measures recent job separations and hiring conditions. GK include neither margin. Their extensions to the baseline add interest rates one at a time, not labor variables. A monetary tightening can weaken demand, raise financing costs, and tighten working capital, and firms can respond on the hours margin.

cores, although the computations were serial. The complete model-construction process across all values of λ required 290 seconds of wall-clock time. The OOS selection based on 2,000 posterior draws for each model required 840 seconds (14 minutes). Increasing the number of posterior draws did not alter the results.

¹⁵In each 7-variable system, the responses of the six core variables are essentially unchanged from the core system in Figure 7. We therefore report only the response of the added variable.

The evidence is mixed. Short-duration unemployment shows no clear response; its posterior median is small and its bands are wide. Hours in both sectors decline, with the largest reductions in the first year. The 68% bands lie below zero over the medium run, while the 90% bands include zero. A monetary tightening therefore puts downward pressures on hours worked. GK measure activity through industrial production alone. These labor margins complement that measure and respond to the same contractionary shock. This group also shows why selection and response strength are distinct. Short-duration unemployment is selected for its predictive content, yet its own response to the identified shock is weak.

Housing activity. The selected system adds three regional housing series: starts in the Northeast and Midwest and permits in the Northeast (Figure 9). This group extends the mortgage-spread channel already in the GK core. The mortgage spread measures the price of housing finance. Permits and starts measure the quantity response, at the planning and construction stages. Housing is among the most interest-sensitive sectors of the economy, and the sector where the policy rate, long rates, mortgage spreads, collateral values, and construction decisions meet. The regional labels carry no structural interpretation. These indicators are selected for their informational content in the model-construction procedure. The selection of more than one regional series indicates that regional housing adjustments are not synchronized and that this heterogeneity carries predictive content for the fitted composite disturbance in the policy equation.

Housing starts and permits have negative posterior medians. The 68% credible band for housing permits includes zero. For housing starts, the 68% credible bands lie strictly below zero over the one-year horizon in the Midwest and the two-year horizon in the Northeast, whereas the 90% bands include zero. A monetary tightening therefore puts downward pressures on residential construction. These quantity responses complement the positive mortgage-spread response in the core system: tighter housing finance is followed by weaker building. The selected additions show that the quantity side of the same channel belongs in the information set.

External and oil markets. The selected system adds the U.S.–U.K. exchange rate and the oil price (Figure 10). Both contain fast-moving information about global financial conditions and cost pressures. The exchange rate reflects differences in monetary conditions across countries and changes in global risk. Oil prices respond to global demand and supply and also affect inflation. In the selected system, the exchange rate helps distinguish domestic policy movements from changes in foreign monetary conditions and global risk, while the oil price helps distinguish policy surprises from global demand and oil supply disturbances. Together, these variables help prevent external and commodity-market disturbances from contaminating the monetary policy shock.

Other U.S. dollar exchange rates are highly correlated with this bilateral rate. As with the regional housing series, the bilateral exchange rate is an indicator selected by the model-construction procedure, not a claim that the United Kingdom has a unique role in U.S. policy transmission. The exchange rate declines persistently. Because the series is measured in U.S. dollars per U.K. pound, the decline is a dollar appreciation after the monetary policy tightening. Its 68% band lies below zero over the medium run, and the upper edge of the 90% band is only barely above zero, so the posterior places overwhelming probability on a decline. The oil price falls on impact and then reverts, but its bands are wide throughout. A monetary tightening therefore puts downward pressures on the oil price, consistent with weaker global demand, though the oil response is only suggestive. GK include no exchange rate or commodity price. These margins condition the system on global conditions that move alongside policy surprises.

Equity valuation and volatility. The selected system adds the price-earnings ratio and the VIX (Figure 11). The two are related but distinct. The price-earnings ratio is a valuation ratio that reflects discount rates, expected earnings growth, and risk compensation. The VIX is a forward-looking measure of expected volatility. These are the first noncore margins credible at the 90% level. The price-earnings ratio falls sharply during the first year, with both bands below zero over the early horizons. Its median later turns positive, but its 68% credible band includes zero. The VIX rises immediately and then decays, with both bands above zero over the early horizons. A monetary policy tightening therefore operates through both valuation and volatility. Lower valuations and higher volatility affect investment, collateral values, risk-bearing capacity, and the willingness of intermediaries to supply credit. GK measure neither margin. These responses show a rapid deterioration in market risk conditions that the core credit spreads do not summarize.

Credit and liquidity conditions. The selected system retains the three core credit-cost measures and adds the Baa-minus-federal-funds (BAA-FF) spread, the TED spread, and the GZ corporate bond spread (Figure 12). These cover different parts of the credit system. The BAA-FF spread combines the transmission of the policy rate with term and corporate-risk components. The TED spread measures the premium on unsecured wholesale dollar funding relative to safe Treasury bills, and captures funding stress and flight to safety. The GZ spread measures the total corporate bond spread. Together with the mortgage and commercial paper spreads, these variables trace credit conditions from borrowers back to the funding markets that support lending.

Across the three core-plus-one reporting systems, the GZ spread rises sharply and is credible at the 90% level over the early horizons, the strongest response of the three. The TED spread rises on impact and is credible at the 90% level at the short end, then decays quickly. The BAA-FF response is hump-shaped and imprecise, not even credible at the 68% level in the medium run. Together, these reporting systems extend GK's central result across

corporate bond pricing, wholesale bank funding, and short-term business credit. The credit channel is not confined to a single spread.

Expected default risk. The GZ spread and the excess bond premium are both in the selected information set, and their difference isolates the spread component associated with expected default risk (Figure 13). GK emphasize that the excess bond premium removes the expected-default component and isolates the part of corporate spreads driven by risk-bearing capacity. The difference $GZS - EBP$ rises sharply after the shock, and both the 68% and 90% bands exclude zero over the early horizons. Monetary tightening therefore widens the corporate bond spread through two channels: the excess bond premium that GK emphasize and the expected-default component that they remove. This is the sharpest extension of their central result. The total rise in corporate borrowing costs reflects both a fall in risk-bearing capacity and a deterioration in the expected creditworthiness of borrowers. Expected default, therefore, is an independently essential part of the credit channel.

Summary. Taken together, the six margins convert GK’s researcher-chosen sequence of interest-rate extensions into one selected information set of economically distinct margins. The selection adds no dense ladder of Treasury yields. It adds labor adjustment, housing variables, equity valuation, volatility, corporate credit, wholesale liquidity, the exchange rate, and the oil price. The strongest evidence lies in the financial margins: the responses of the price-earnings ratio, VIX, GZ spread, TED spread, and expected-default risk are credible at the 90% level over the early horizons. The BAA–FF response is imprecise. The real activity and external margins show downward pressures with less precision.

The expected-default result extends GK’s central finding most directly. In the core system augmented with the selected GZ spread, the component they remove from the corporate spread responds materially to monetary policy. The selected information set is therefore a disciplined enlargement of GK. It preserves their external-instrument strategy and emphasis on credit costs, replaces a single anchor and one instrument with anchor-free joint estimation from multiple instruments, and replaces researcher-chosen model expansion with iterative construction and OOS selection. The selected information set is larger than the GK baseline, narrower than the full candidate data set, and organized around distinct margins that trace monetary policy transmission through financial conditions.

VIII. CONCLUSION

SVAR evidence is conditional on an information set. When hundreds of candidate series are available, the surrounding information set cannot be chosen credibly by hand. This paper formalizes that choice through an algorithm-driven Bayesian procedure. The researcher specifies the economic question, the core variables, and the identifying restrictions. For each value of the model complexity parameter, the construction step admits remaining candidates with predictive content for the fitted composite disturbances of the current system,

reestimates the enlarged SVAR, and iterates to a terminal information set. For each loss, the Bayesian OOS criterion retains systems credibly better than the reference; when none exist, it retains systems credibly indistinguishable from the reference. The final rule selects the largest system in the union of the retained sets under squared and absolute loss. We establish finite termination, uniqueness of the construction path and terminal system, an exact stopping characterization, invariance of the auxiliary admission rule at a given iteration to the measurement units of the remaining candidates, stability under small numerical perturbations, and existence and uniqueness of the OOS-selected system.

The two applications show that the information set can alter the central economic conclusion. In the household credit application, the procedure selects four measures of housing production that neither MSV nor BPSS include. A household credit shock raises household credit while housing production falls and the posterior median of output shows no initial boom. A separate housing production shock raises construction, output, and household credit together. Within the selected system, output rises with expanding residential construction, not with household credit alone. Models of household credit and the business cycle should therefore allow housing production to move separately from household credit.

In the monetary policy application, joint estimation with three instruments and no anchor strengthens GK’s central credit spread channel. In the core system augmented with the selected GZ spread, monetary policy tightening raises both the excess bond premium and expected default risk. The application delivers an independent theoretical result. Under the maintained proxy restrictions, a vector of external instruments identifies one shock and its equation if and only if at least one instrument is correlated with that shock. The theorem allows any number of instruments for one shock, extends to several shocks and their corresponding instrument vectors under the joint covariance-admissibility condition, and gives explicit formulas for the coefficients of the identified equation and its impulse responses without a reduced-form innovation anchor.

The framework is designed to be modular. The auxiliary optimization is one implementation of the construction step; alternative optimization criteria could use different regularizers, group a variable with its lags, or target several equations with different tuning parameters. The Bayesian selection step could use other loss functions or proper scoring rules for the full predictive distribution, and the candidate set could include constructed latent factors along with observed variables. The architecture, however, remains the same: the researcher defines the economic question and identifying restrictions, the algorithm constructs candidate information sets, Bayesian estimation incorporates parameter uncertainty, and OOS evidence selects model complexity. The specific theoretical guarantees and empirical performance would have to be established for each alternative. Our theoretical results provide those guarantees for the procedure studied here and establish a well-defined, reproducible, tractable, and economically informative benchmark for evaluating such alternatives.

The paper offers a broader lesson. Identification restrictions are stated, defended, and debated; yet the variables on which they operate are often chosen informally. When the set of candidate variables is vast, this asymmetry is no longer defensible. An explicit algorithm makes the information set reproducible and open to inspection, and the applications show that this discipline can alter the economic interpretation of an identified shock. Identification has long been disciplined. In the age of big data, the information set deserves the same discipline. This paper provides it.

TABLE 1. Variables in the selected household credit system

Short	Original ticker	Economic description
IP	INDPRO	Industrial production index.
BC	BUSLOANS	Commercial and industrial loans.
HHC	HH	Consumer and real estate loans at all commercial banks.
HRS-G	CES0600000007	Average weekly hours in goods-producing industries.
HS-S	HOUSTS	Housing starts in the South.
PER-NE	PERMITNE	New private housing permits in the Northeast, at a seasonally adjusted annual rate.
T1Y-FF	T1YFFM	One-year Treasury rate minus the federal funds rate.
EBP	EBP_OA	Option-adjusted excess bond premium of GZ—the corporate bond spread component not attributable to expected default risk.
GZS	GZ_SPR	GZ corporate bond spread, comprising the component associated with expected default risk and the excess bond premium.
HRS-M	AWHMAN	Average weekly hours in manufacturing.
PER	PERMIT	New private housing permits, at a seasonally adjusted annual rate.
UI	CLAIMSx	Initial claims for unemployment insurance.
PER-W	PERMITW	New private housing permits in the West, at a seasonally adjusted annual rate.

Note: The table reports the 13 variables included in the selected system at $\lambda = 0.1129$. It gives the short abbreviations used in the discussion in the first column, the original series names in the second column, and their economic meanings in the third column. The original tickers are FRED series IDs whenever available.

Variables entered in log levels are reported as approximate percentage changes; rates, spreads, and shares are reported in percentage points.

TABLE 2. Variables in the selected monetary policy system

Short	Original ticker	Economic description
T1Y	GS1	One-year Treasury rate.
CPI	CPIAUCSL	Consumer price index, all items.
IP	INDPRO	Industrial production index.
EBP	EBP_OA	GZ option-adjusted excess bond premium, the corporate bond spread component not attributable to expected default risk.
MGS	MortgageSpliceGS10	Mortgage spread relative to the ten-year Treasury rate.
CPS	CP3M_TB3MS	Three-month commercial paper spread relative to the three-month Treasury bill rate.
U5	UEMPLT5	Civilians unemployed for less than 5 weeks.
HRS-G	CES0600000007	Average weekly hours in goods-producing industries.
HRS-M	AWHMAN	Average weekly hours in manufacturing.
HS-NE	HOUSTNE	Housing starts in the Northeast.
HS-MW	HOUSTMW	Housing starts in the Midwest.
PER-NE	PERMITNE	New private housing permits in the Northeast, at a seasonally adjusted annual rate.
PE	S&P PE ratio	Price-earnings ratio for the S&P composite common stock index.
BAA-FF	BAAFFM	Moody's Baa corporate bond yield minus the federal funds rate.
FX-UK	EXUSUKx	U.S.–U.K. foreign exchange rate.
OIL	OILPRICEx	Crude oil price, spliced WTI and Cushing series.
VIX	VIXCLSx	CBOE implied volatility index.
TED	MED3_TB3MS	Three-month Eurodollar deposit rate minus the three-month Treasury bill rate.
GZS	GZ_SPR	GZ corporate bond spread, comprising the component associated with expected default risk and the excess bond premium.

Note: The table reports the variables included in the selected monetary policy system. It gives the short abbreviations used in the discussion in the first column, the original series names in the second column, and their economic meanings in the third column. The original tickers are FRED series IDs whenever available.

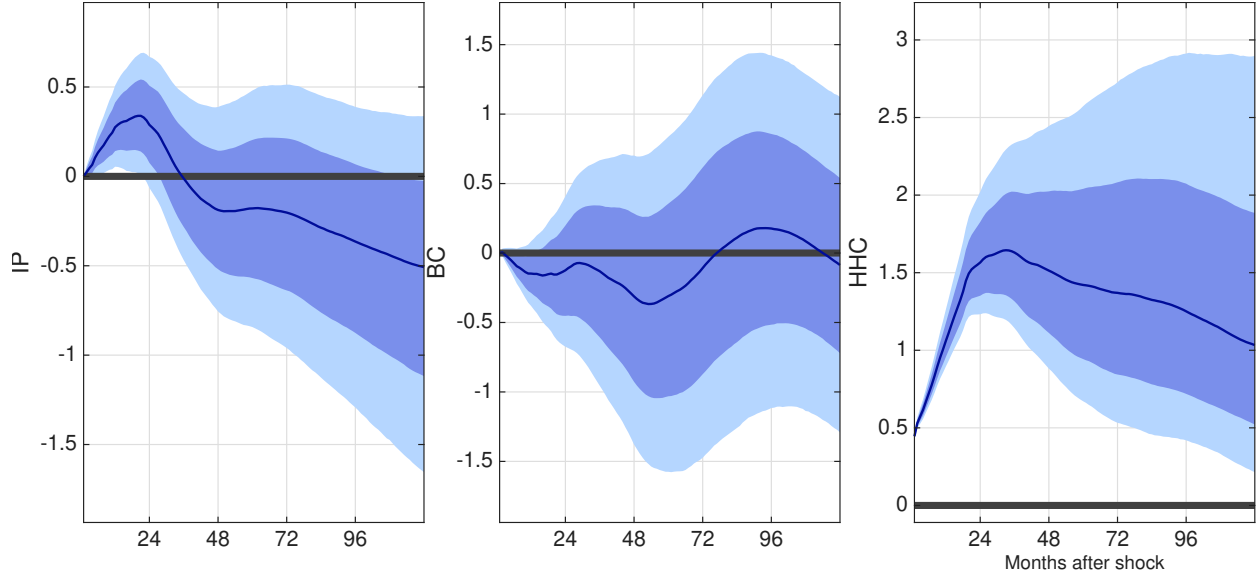


FIGURE 1. Household credit shock and its dynamic impacts. This figure replicates MSV’s Figure I using updated monthly data and the monthly Sims–Zha prior. All responses are expressed in percent. Unless otherwise stated, throughout this paper we follow GK and use “percent” as shorthand: variables entered in log levels are reported as approximate percentage changes, whereas rates, spreads, and shares are reported in percentage points. The dark band shows the 68% posterior credible band, and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

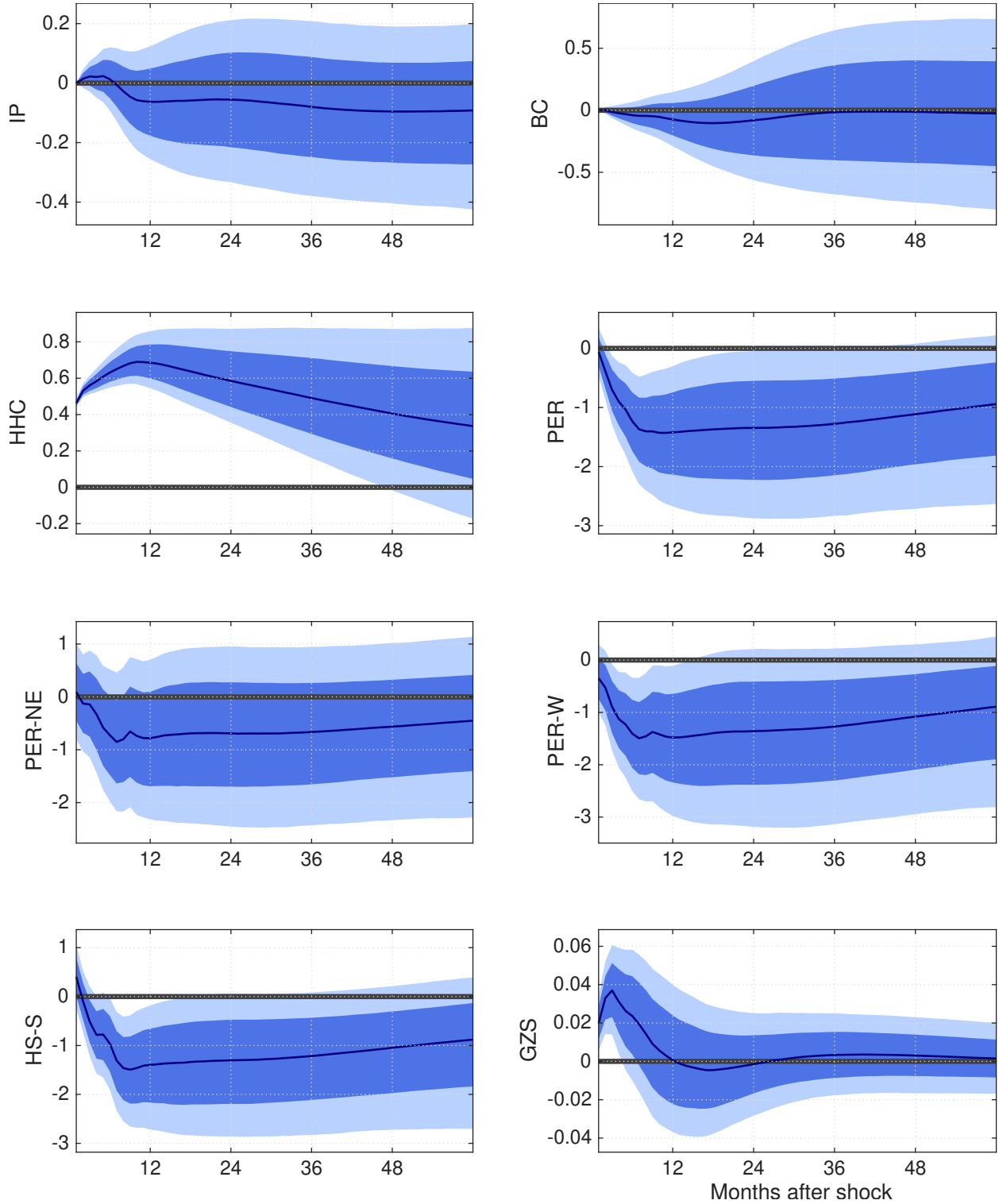


FIGURE 2. Household credit shock and its dynamic effects. The impulse responses are estimated from the selected 13-variable system under the recursive identification used by MSV. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

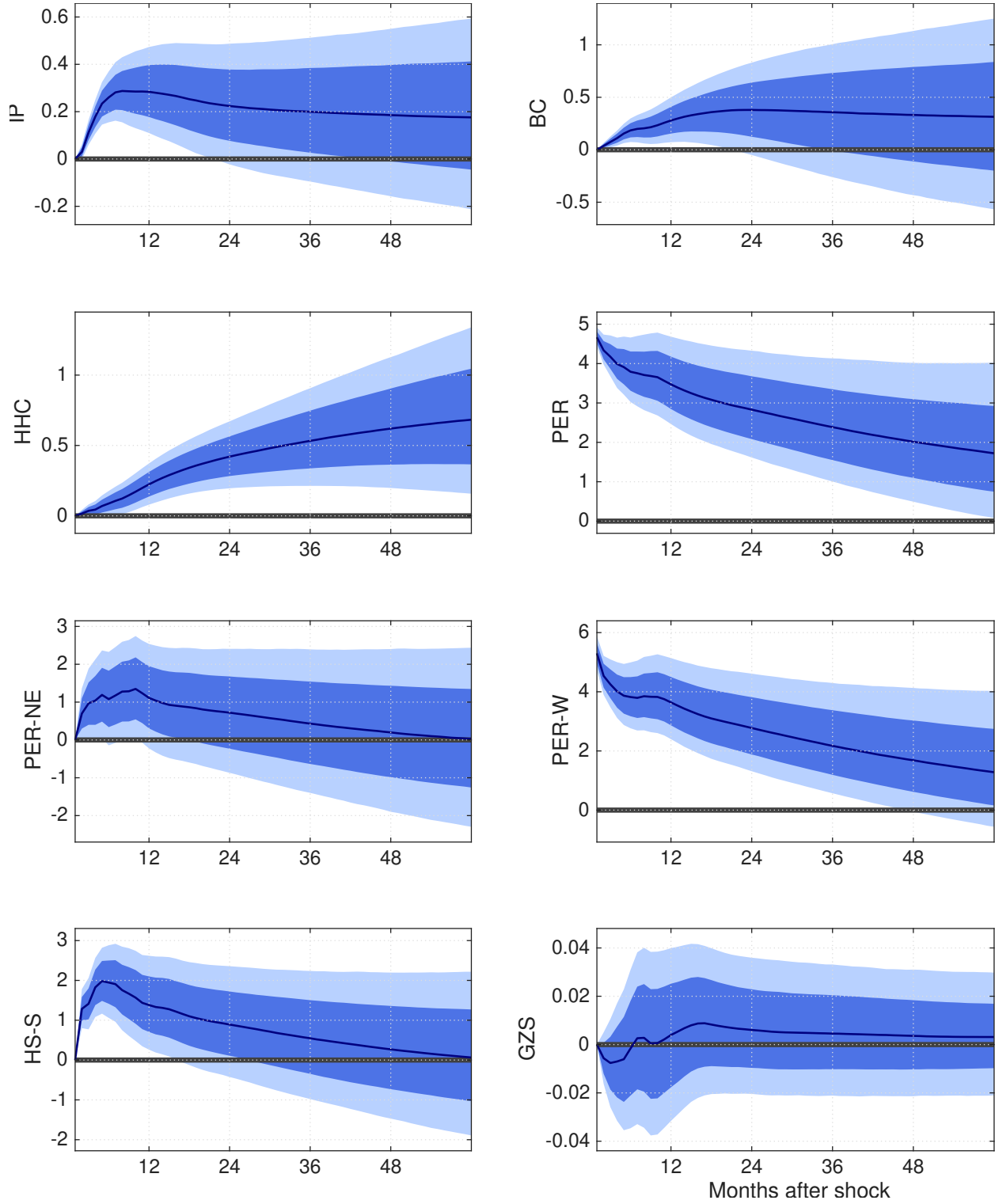


FIGURE 3. Housing production shock and its dynamic impacts. The impulse responses are estimated from the selected 13-variable system under the recursive identification used by MSV. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band, and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

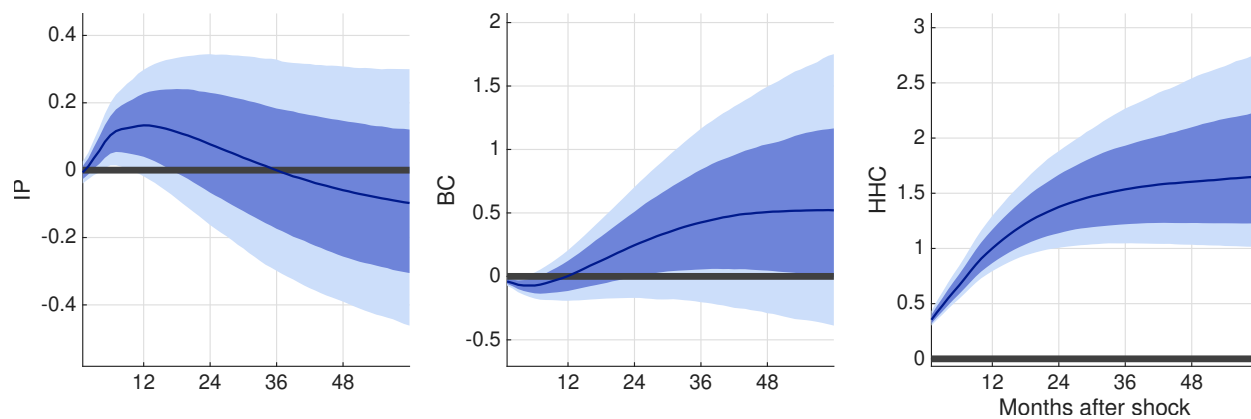


FIGURE 4. Household credit shock and its dynamic impacts. This figure replicates the IP, BC, and HHC responses in the second column of BPSS's Figure 2 using updated monthly data and the monthly Sims–Zha prior. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band, and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

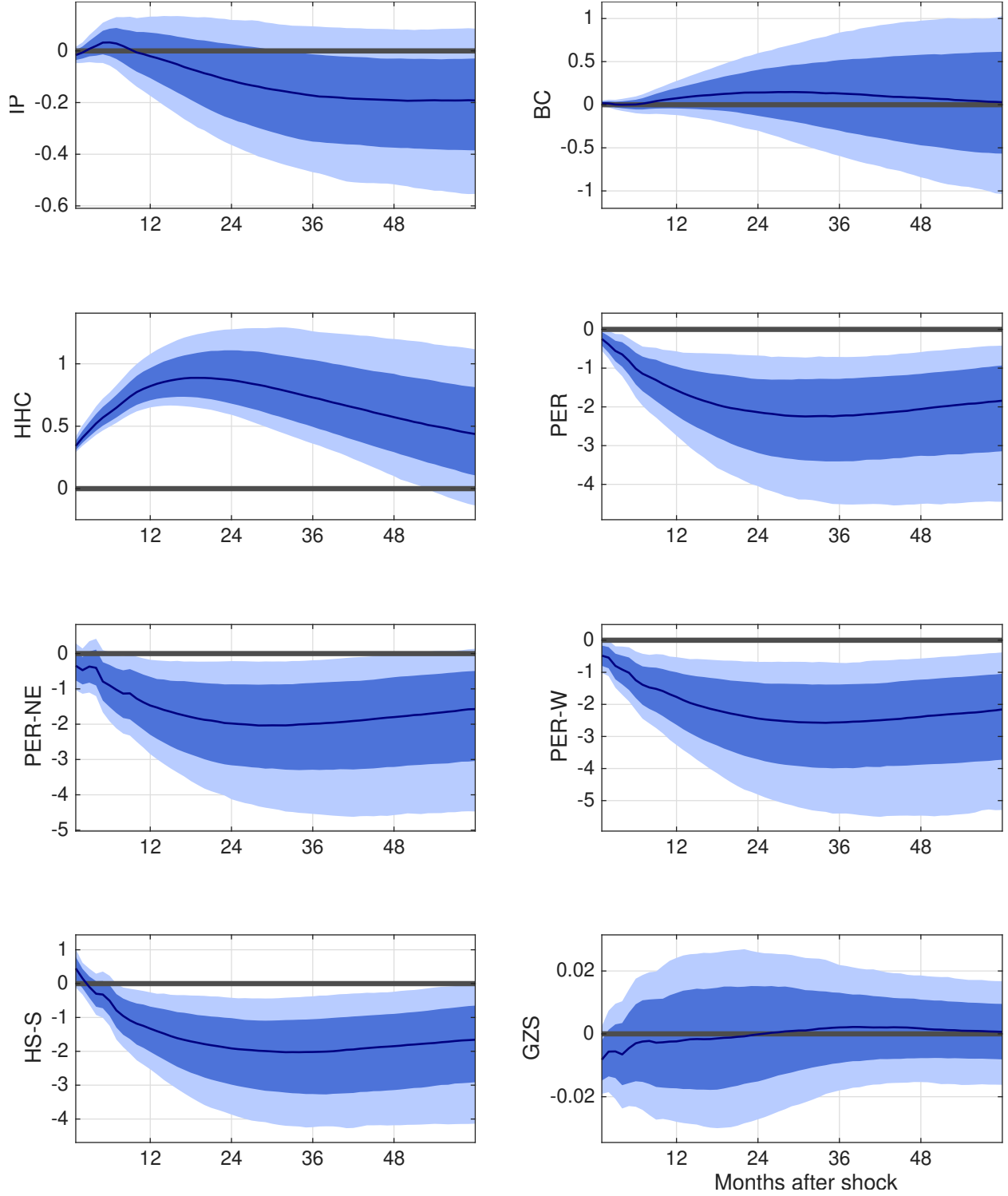


FIGURE 5. Household credit shock and its dynamic effects. The impulse responses are estimated from the selected 13-variable system using the heteroskedasticity-based identification of BPSS. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

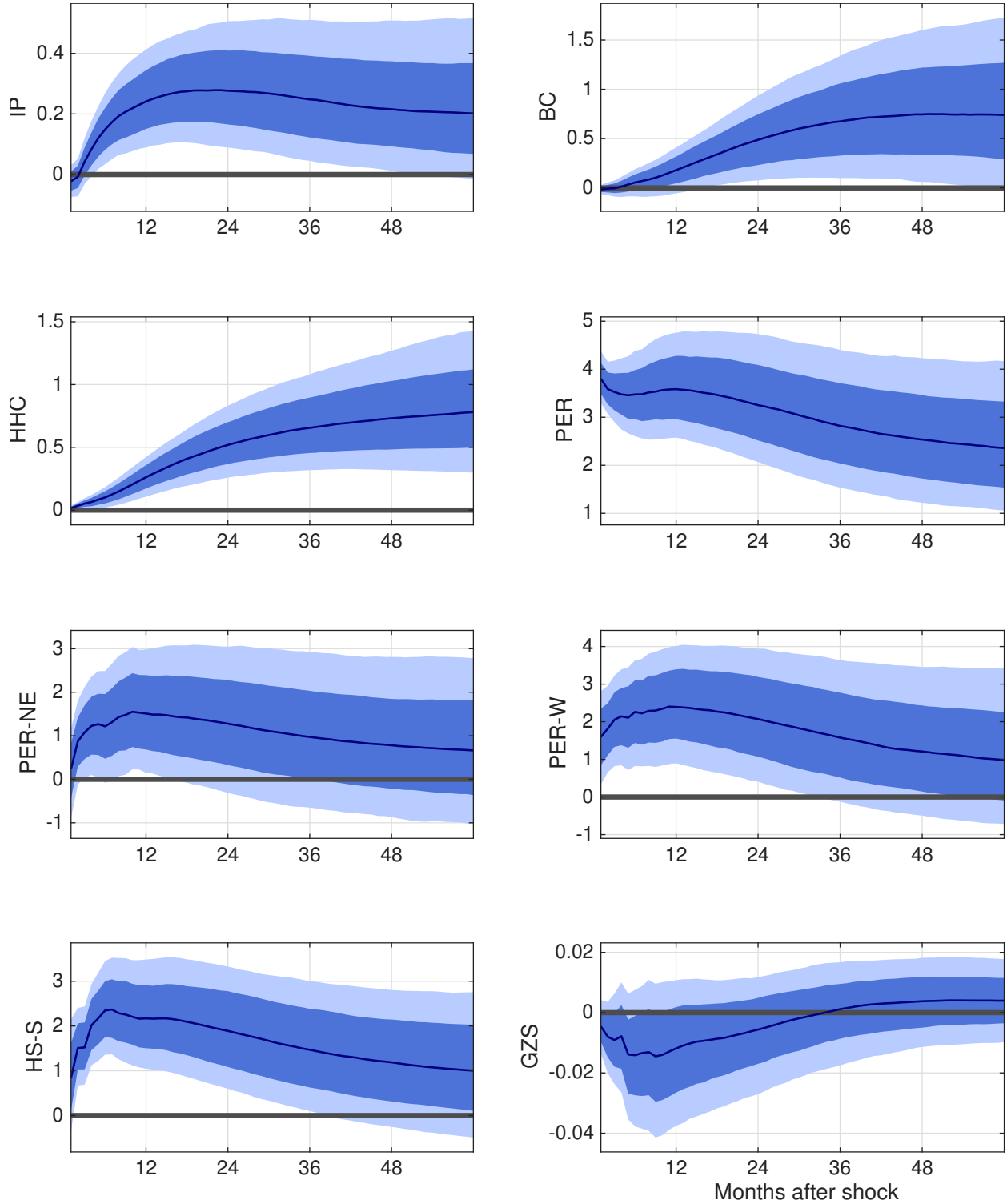


FIGURE 6. Housing production shock and its dynamic impacts. The impulse responses are estimated from the selected 13-variable system using the heteroskedasticity-based identification of BPSS. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band, and the light band the 90% posterior credible band. See Table 1 for complete variable descriptions.

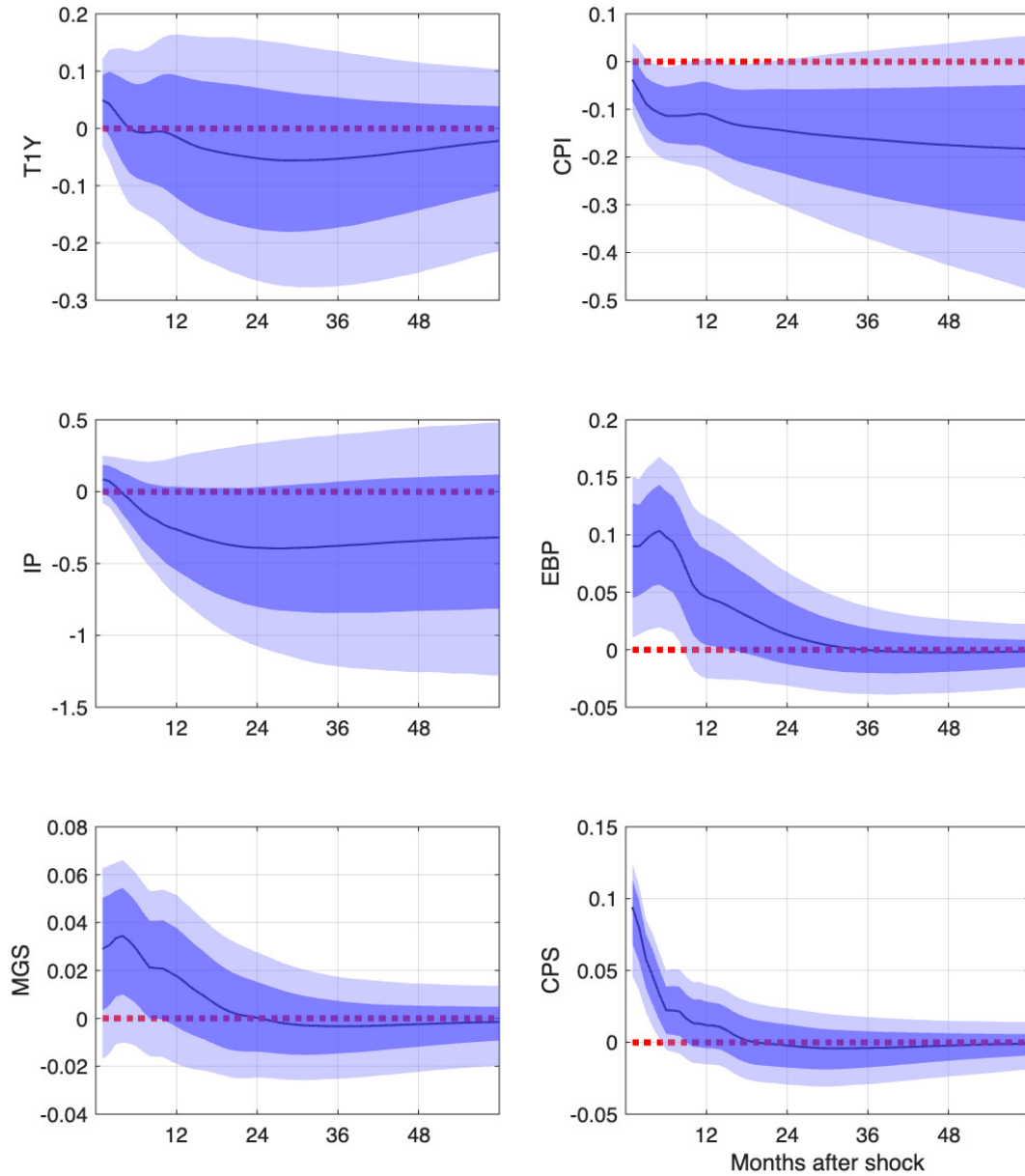


FIGURE 7. Monetary policy shock and its transmission through the core variables. Following GK, all dynamic responses are expressed in percent. “T1Y” denotes the one-year Treasury rate, “CPI” the consumer price index, “IP” industrial production, “EBP” the excess bond premium, “MGS” the mortgage spread, and “CPS” the commercial paper spread. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

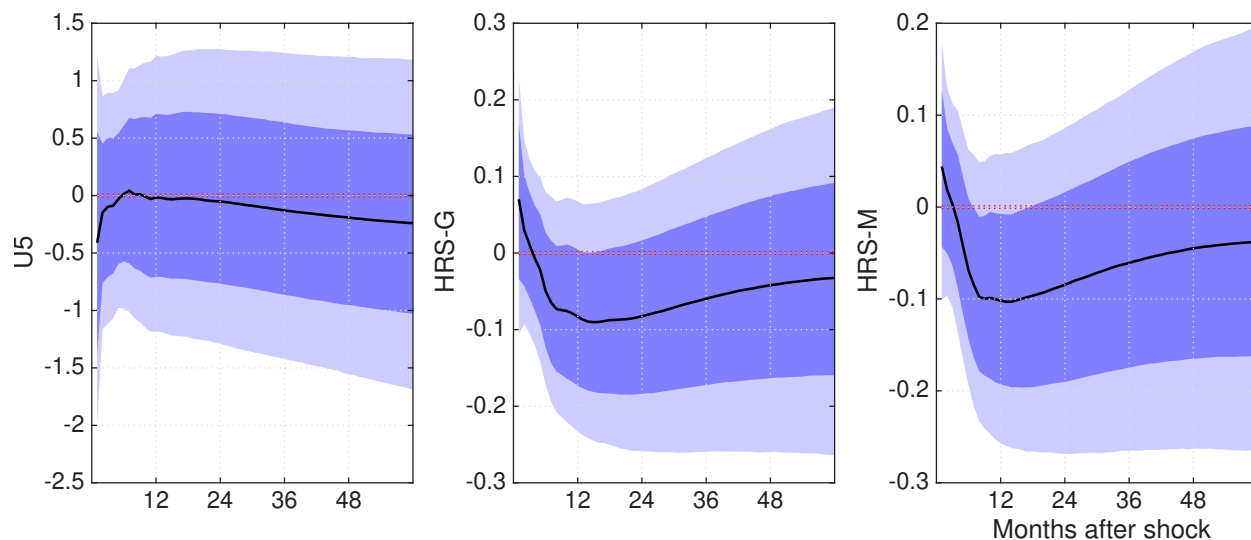


FIGURE 8. Monetary policy shock and its transmission through selected labor market variables. Following GK, all dynamic responses are expressed in percent. “U5” denotes unemployed for less than 5 weeks, “HRS-G” hours in goods-producing industries, and “HRS-M” hours in manufacturing. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

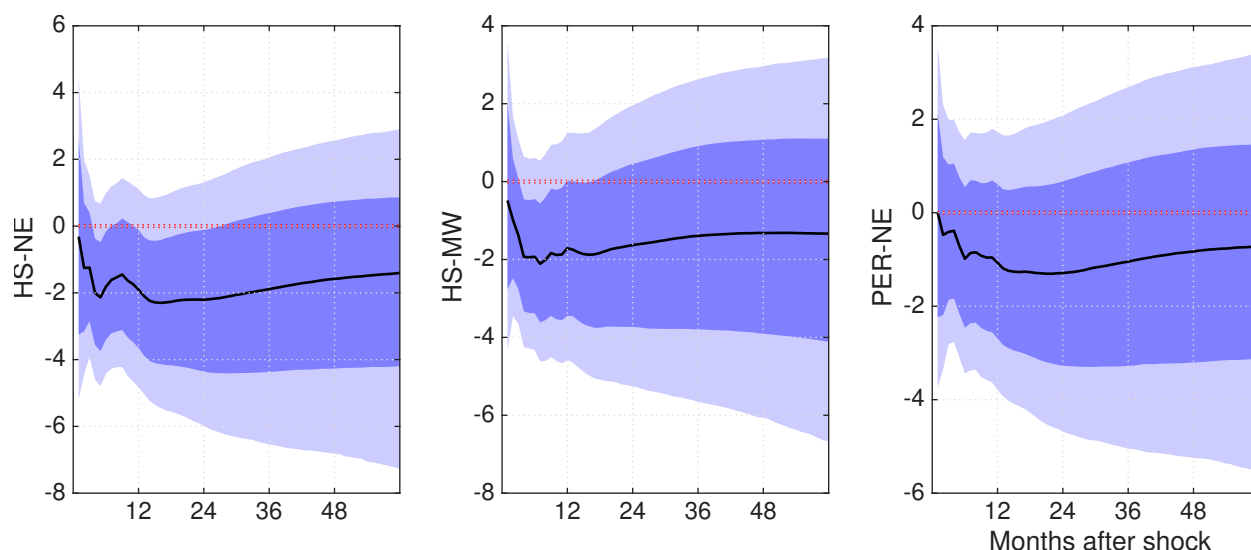


FIGURE 9. Monetary policy shock and its transmission through selected housing variables. Following GK, all dynamic responses are expressed in percent. “HS-NE” denotes housing starts in the Northeast, “HS-MW” housing starts in the Midwest, and “PER-NE” new housing permits in the Northeast. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

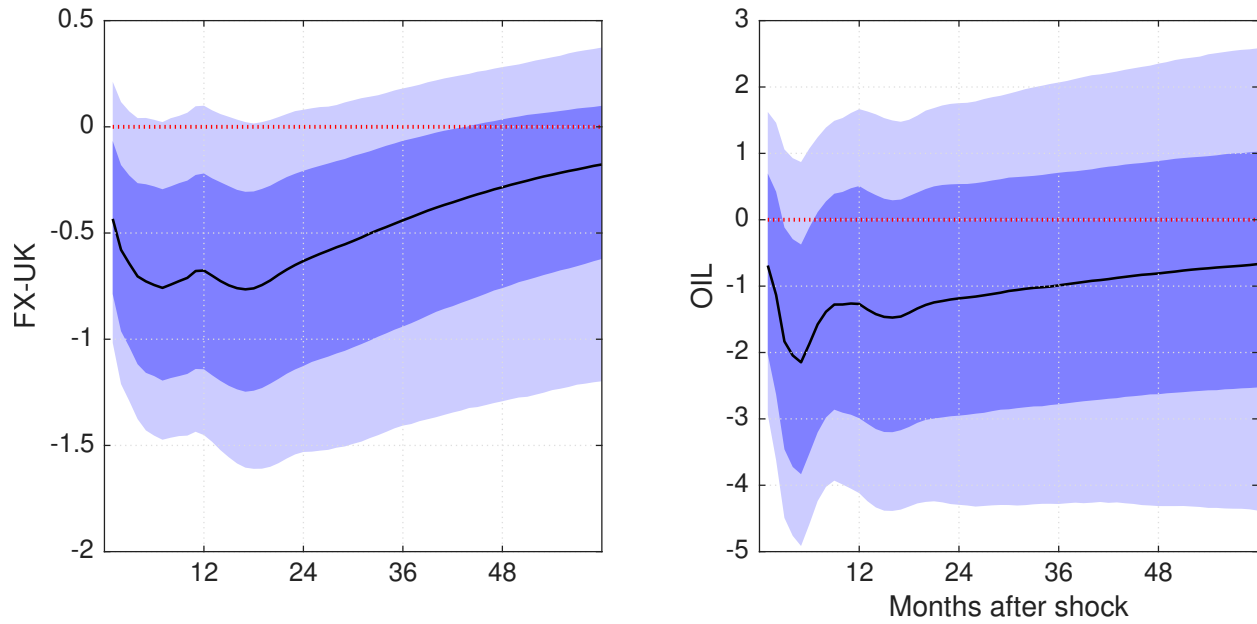


FIGURE 10. Monetary policy shock and its transmission through selected external and commodity variables. Following GK, all dynamic responses are expressed in percent. “FX-UK” denotes the U.S./U.K. foreign exchange rate, and “OIL” denotes the oil price. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

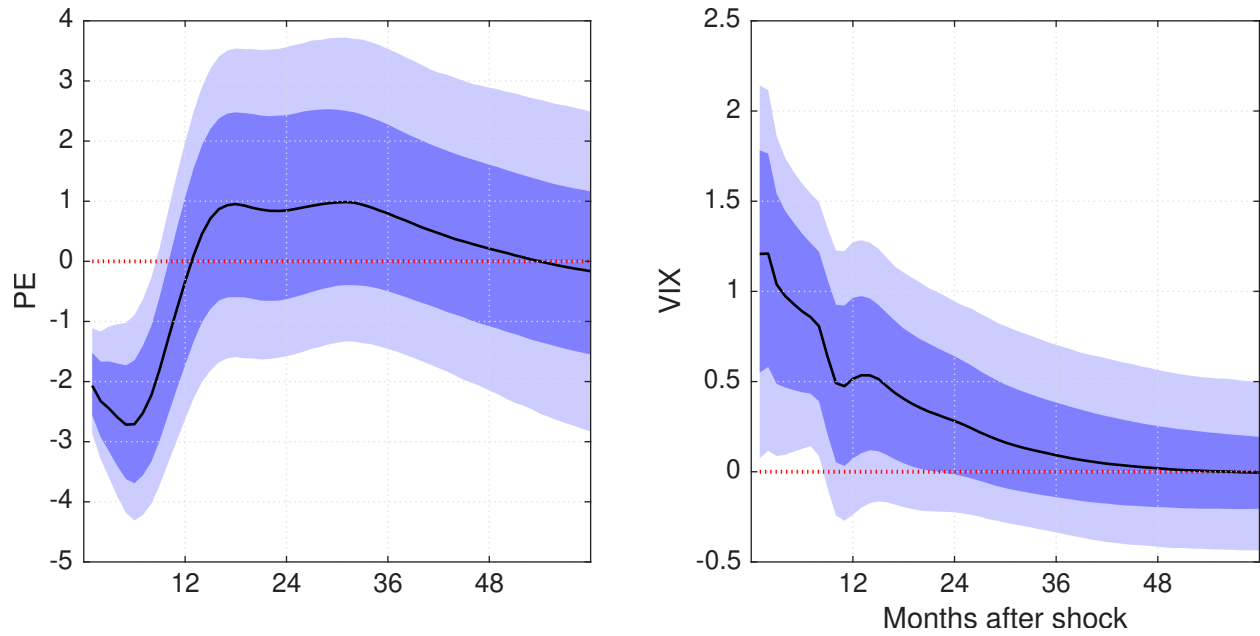


FIGURE 11. Monetary policy shock and its transmission through selected equity and volatility variables. Following GK, all dynamic responses are expressed in percent. “PE” denotes the price-earnings ratio, and “VIX” denotes stock market volatility. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

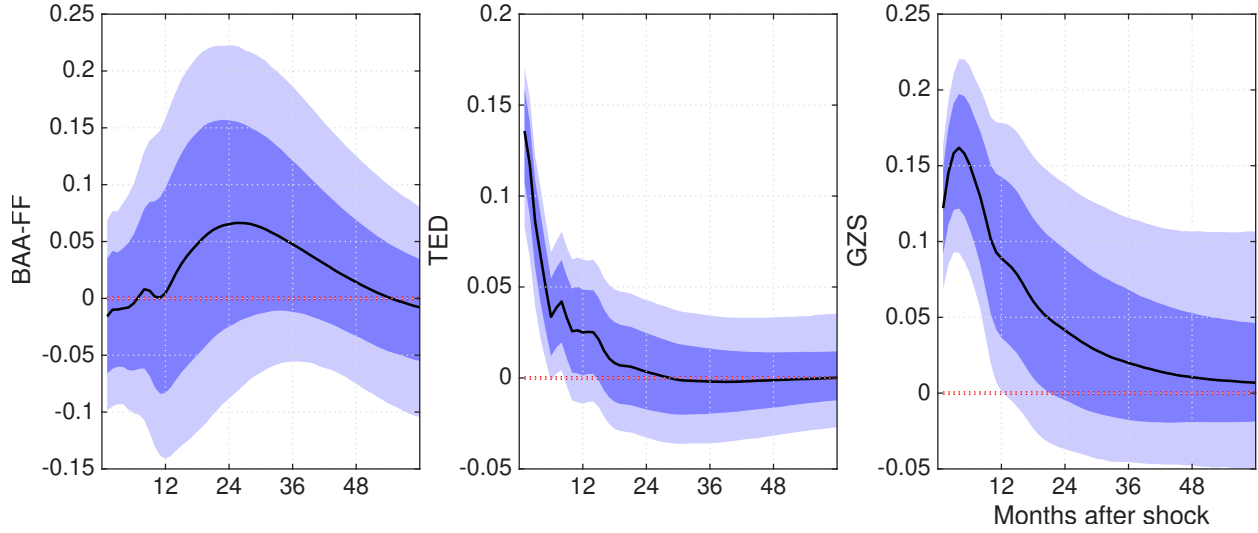


FIGURE 12. Monetary policy shock and its transmission through selected credit and liquidity variables. Following GK, all dynamic responses are expressed in percent. “BAA-FF” denotes the Baa corporate bond yield spread over the federal funds rate, “TED” the Eurodollar deposit rate spread over the three-month Treasury bill rate, and “GZS” the GZ corporate bond spread. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. See Table 2 for complete variable descriptions.

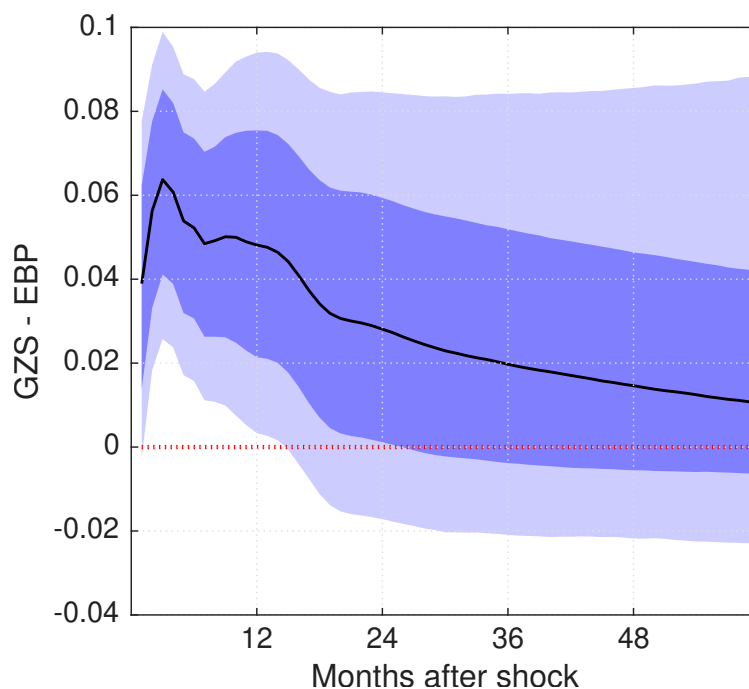


FIGURE 13. Monetary policy shock and its transmission through expected default risk. The responses are estimated from the six-variable core system augmented with the selected GZ spread. “EBP” denotes the excess bond premium, and “GZS” denotes the GZ corporate bond spread. The difference $GZS - EBP$ measures the spread component attributable to expected default risk. All responses are expressed in percentage points. The dark and light bands depict the 68% and 90% posterior credible bands, respectively. See Table 2 for complete variable descriptions.

In this appendix, sections, tables, figures, and equations are labeled with the prefix “S” to denote supplemental material.

APPENDIX S1. PROOFS FOR MODEL CONSTRUCTION AND OOS SELECTION

Proof of Proposition 1. By (14), $\widetilde{\mathcal{J}}^{(k)}(\lambda)$ is a subset of $\widetilde{\mathcal{J}}^{(k)} = \mathcal{N} \setminus \mathcal{J}^{(k)}$. Hence, $\widetilde{\mathcal{J}}^{(k)}(\lambda)$ and $\mathcal{J}^{(k)}$ are disjoint. The update in (15) therefore gives

$$\mathcal{J}^{(k+1)} = \mathcal{J}^{(k)} \cup \widetilde{\mathcal{J}}^{(k)}(\lambda), \quad |\mathcal{J}^{(k+1)}| = |\mathcal{J}^{(k)}| + |\widetilde{\mathcal{J}}^{(k)}(\lambda)|.$$

The first equality implies $\mathcal{J}^{(k)} \subseteq \mathcal{J}^{(k+1)}$. It also implies that an index already contained in $\mathcal{J}^{(k)}$ cannot be removed by any later update. If $\widetilde{\mathcal{J}}^{(k)}(\lambda) \neq \emptyset$, the second equality implies $|\mathcal{J}^{(k+1)}| > |\mathcal{J}^{(k)}|$. Every nonterminal update therefore adds at least one index from the finite set $\mathcal{N} \setminus \mathcal{J}^{(0)}$, which contains $n - n^{(0)}$ indices. There can be no more than $n - n^{(0)}$ strict updates. The terminal index consequently satisfies $\bar{k} \leq n - n^{(0)}$. The procedure evaluates the initial system and then evaluates one enlarged system after each strict update. Hence, the number of current systems estimated along the path is $\bar{k} + 1 \leq n - n^{(0)} + 1$. $\square$

Proof of Proposition 2. The initial index set $\mathcal{J}^{(0)}$ is fixed by the economic application. Suppose that $\mathcal{J}^{(k)}$ is uniquely determined. The corresponding variable vector $y_t^{(k)}$, the remaining index set $\widetilde{\mathcal{J}}^{(k)}$, the index mapping $j \mapsto \iota_j^{(k)}$, and the standardized regressor vector $x_{t,\text{std}}^{(k)}$ are then uniquely determined because the standardization moments are computed once and held fixed.

Assumption 1 ensures that the argmax set in (8) is nonempty, and the prespecified deterministic rule selects one posterior mode $\widehat{\theta}^{(k)}$ and hence one fitted composite disturbance vector $\widehat{u}_t^{(k)}$. The same assumption determines one minimizer of (11) for every $q \in \mathcal{Q}$. The coefficient blocks in (13), the selected set $\widetilde{\mathcal{J}}^{(k)}(\lambda)$ in (14), and the updated set $\mathcal{J}^{(k+1)}$ in (15) are consequently unique. Induction from $k = 0$ establishes uniqueness of the entire path. Proposition 1 establishes finite termination, and the terminal index set and variable vector are therefore unique. $\square$

Proof of Lemma 1. For fixed β_q , the criterion in (11) is a strictly convex quadratic function of d_q . Differentiating with respect to d_q and setting the derivative equal to zero gives

$$-\frac{1}{T_{\mathcal{Q}}} \mathbf{1}' (\widehat{\mathbf{u}}_q^{(k)} - d_q \mathbf{1} - X^{(k)} \beta_q) = 0,$$

which gives the stated expression for $d_q(\beta_q)$. The residual after substitution is

$$\widehat{\mathbf{u}}_q^{(k)} - d_q(\beta_q) \mathbf{1} - X^{(k)} \beta_q = M (\widehat{\mathbf{u}}_q^{(k)} - X^{(k)} \beta_q).$$

The ℓ_1 term does not involve the intercept. Substitution therefore yields (16), and the minimizing coefficient vector is unchanged. $\square$

Proof of Proposition 3. If $p_k = 0$, no candidate remains, $\widetilde{\mathcal{J}}^{(k)}(\lambda) = \emptyset$, and the left-hand side of (17) is zero by convention. The result therefore holds. Suppose $p_k > 0$. By Lemma 1,

the estimated coefficients on the regressors are unchanged when the dependent variable and regressors are centered. For equation q , the auxiliary optimization is therefore equivalent to

$$\min_{\beta_q} \left\{ \frac{1}{2T_{\mathcal{L}}} \|M\hat{\mathbf{u}}_q^{(k)} - MX^{(k)}\beta_q\|_2^2 + \lambda \|\beta_q\|_1 \right\}.$$

The objective is convex. The vector $\beta_q = 0$ is a minimizer if and only if zero belongs to its subdifferential at $\beta_q = 0$. This condition is

$$0 \in -\frac{1}{T_{\mathcal{L}}} (X^{(k)})' M\hat{\mathbf{u}}_q^{(k)} + \lambda \partial \|\beta_q\|_1 \Big|_{\beta_q=0}.$$

Since

$$\partial \|\beta_q\|_1 \Big|_{\beta_q=0} = \{z : \|z\|_{\infty} \leq 1\},$$

the zero vector is a minimizer if and only if

$$\left\| \frac{1}{T_{\mathcal{L}}} (X^{(k)})' M\hat{\mathbf{u}}_q^{(k)} \right\|_{\infty} \leq \lambda.$$

Under Assumption 1, the auxiliary minimizer is unique. Hence, the condition above is equivalent to $\hat{\beta}_q^{(k)} = 0$. By (14), no remaining candidate is admitted if and only if $\hat{\beta}_q^{(k)} = 0$ for every $q \in \mathcal{Q}$. Therefore,

$$\widetilde{\mathcal{S}}^{(k)}(\lambda) = \emptyset$$

if and only if

$$\max_{q \in \mathcal{Q}} \left\| \frac{1}{T_{\mathcal{L}}} (X^{(k)})' M\hat{\mathbf{u}}_q^{(k)} \right\|_{\infty} \leq \lambda.$$

□

Proof of Proposition 4. For a transformed remaining candidate $y_{i,t}^{\text{new}} = \alpha_i + d_i y_{i,t}$, its mean and standard deviation satisfy

$$\bar{y}_i^{\text{new}} = \alpha_i + d_i \bar{y}_i, \quad \hat{\sigma}_i^{\text{new}} = |d_i| \hat{\sigma}_i.$$

Its standardized value is therefore

$$\frac{y_{i,t}^{\text{new}} - \bar{y}_i^{\text{new}}}{\hat{\sigma}_i^{\text{new}}} = \text{sgn}(d_i) y_{i,t}^{\text{std}}.$$

Consequently, the new standardized regressor matrix has the form $X_{\text{new}}^{(k)} = X^{(k)} H$, where H is diagonal, every diagonal entry equals 1 or -1 , and the same sign is repeated for the contemporaneous value and lags of a given variable. The matrix satisfies $H^{-1} = H$ and preserves the ℓ_1 norm. For any coefficient vector β_q , define $\beta_{q,\text{new}} = H\beta_q$. Then

$$X_{\text{new}}^{(k)} \beta_{q,\text{new}} = X^{(k)} H H \beta_q = X^{(k)} \beta_q, \quad \|\beta_{q,\text{new}}\|_1 = \|\beta_q\|_1.$$

The original and transformed auxiliary criteria consequently have the same values under this one to one mapping. Uniqueness in Assumption 1 implies that their minimizers are related by $\hat{\beta}_{q,\text{new}} = H\hat{\beta}_q$. Multiplication by H changes only coefficient signs. Every coefficient block

is zero before the transformation if and only if the corresponding block is zero after the transformation. Equation (14) therefore gives the same selected variable indices. $\square$

Proof of Proposition 5. Assumption 1 ensures that the argmax set in (8) is nonempty and that the auxiliary minimizers $\widehat{\beta}_q^{(k)}$ are unique for every $q \in \mathcal{Q}$. The prespecified deterministic rule selects one posterior mode $\widehat{\theta}^{(k)}$ and hence one fitted composite disturbance vector. Each nonzero coefficient index set $\mathcal{A}_q^{(k)}$ is therefore uniquely determined before the perturbation. Assumption 2 provides the conditions under which these index sets are preserved under sufficiently small changes in the fitted composite disturbances.

If $p_k = 0$, no regressor remains in the auxiliary optimization. Every nonzero coefficient index set and the selected variable set are empty, and the next index set is unchanged under any perturbation. The conclusion holds for any $\epsilon_k > 0$. Suppose $p_k > 0$.

Fix an equation q and write

$$A = \mathcal{A}_q^{(k)}.$$

First suppose that A is nonempty. Let

$$s_A = \text{sgn}(\widehat{\beta}_{q,A}^{(k)})$$

and define

$$Q_A = \frac{1}{T_{\mathcal{L}}}(\widetilde{X}_A^{(k)})' \widetilde{X}_A^{(k)}.$$

Assumption 2 makes Q_A positive definite and therefore invertible. The first order condition for the original auxiliary problem on the nonzero coordinates is

$$\frac{1}{T_{\mathcal{L}}}(\widetilde{X}_A^{(k)})' \left(\widetilde{\mathbf{u}}_q^{(k)} - \widetilde{X}_A^{(k)} \widehat{\beta}_{q,A}^{(k)} \right) = \lambda s_A.$$

Replace $\widehat{\mathbf{u}}_q^{(k)}$ by $\widehat{\mathbf{u}}_q^{(k)} + h_q$. After centering, the dependent variable becomes $\widetilde{\mathbf{u}}_q^{(k)} + Mh_q$. Holding the coefficient support equal to A and the sign vector equal to s_A , the first order condition for the perturbed problem is

$$\frac{1}{T_{\mathcal{L}}}(\widetilde{X}_A^{(k)})' \left(\widetilde{\mathbf{u}}_q^{(k)} + Mh_q - \widetilde{X}_A^{(k)} \beta_A(h_q) \right) = \lambda s_A.$$

Subtracting the original first order condition and using the definition of Q_A gives

$$\beta_A(h_q) = \widehat{\beta}_{q,A}^{(k)} + Q_A^{-1} \frac{1}{T_{\mathcal{L}}}(\widetilde{X}_A^{(k)})' Mh_q. \quad (\text{S1})$$

The mapping $h_q \mapsto \beta_A(h_q)$ is linear and continuous, and

$$\beta_A(0) = \widehat{\beta}_{q,A}^{(k)}.$$

Assumption 2 gives

$$\min_{j \in A} \left| \widehat{\beta}_{q,j}^{(k)} \right| \geq b_q > 0.$$

Consequently, there exists $\epsilon_{q,1} > 0$ such that

$$\|h_q\|_2 < \epsilon_{q,1}$$

implies

$$\operatorname{sgn}(\beta_A(h_q)) = s_A.$$

Thus, every coefficient indexed by A remains nonzero and retains its original sign.

Define the residual associated with (S1) by

$$\mathbf{v}_q(h_q) = \tilde{\mathbf{u}}_q^{(k)} + Mh_q - \tilde{X}_A^{(k)}\beta_A(h_q).$$

If A^c is nonempty, define for each index $j \notin A$ the dual quantity

$$d_j(h_q) = \frac{1}{T_{\mathcal{L}}}(\tilde{X}_j^{(k)})'\mathbf{v}_q(h_q).$$

Equation (S1) implies that each mapping $h_q \mapsto d_j(h_q)$ is continuous. Assumption 2 gives the strict inactive-coordinate bound

$$|d_j(0)| \leq \lambda - \delta_q \quad \text{for every } j \notin A,$$

where $\delta_q > 0$. Because there are finitely many inactive coordinates, there exists $\epsilon_{q,2} > 0$ such that

$$\|h_q\|_2 < \epsilon_{q,2}$$

implies

$$|d_j(h_q)| < \lambda \quad \text{for every } j \notin A.$$

If A^c is empty, set $\epsilon_{q,2} = +\infty$.

Let

$$\epsilon_q^A = \min\{\epsilon_{q,1}, \epsilon_{q,2}\}.$$

When $\|h_q\|_2 < \epsilon_q^A$, the coefficient vector whose entries equal $\beta_A(h_q)$ on A and zero outside A satisfies the complete optimality conditions for the perturbed auxiliary problem. On A , the first order condition holds with sign vector s_A . Outside A , the strict subgradient inequalities hold.

It remains to establish that this perturbed minimizer is unique. Let $\beta(h_q)$ denote the coefficient vector just constructed, and define

$$z(h_q) = \frac{1}{\lambda T_{\mathcal{L}}}(\tilde{X}^{(k)})'\mathbf{v}_q(h_q).$$

The optimality conditions give

$$z_A(h_q) = s_A.$$

When A^c is nonempty, they also give

$$\|z_{A^c}(h_q)\|_{\infty} < 1.$$

For any coefficient vector β , convexity of the squared-error term and the subgradient inequality for the ℓ_1 norm imply

$$\begin{aligned} \mathcal{L}_q(\beta; h_q) - \mathcal{L}_q(\beta(h_q); h_q) &= \frac{1}{2T_{\mathcal{L}}} \left\| \tilde{X}^{(k)} (\beta - \beta(h_q)) \right\|_2^2 \\ &\quad + \lambda \left[\|\beta\|_1 - \|\beta(h_q)\|_1 - z(h_q)' (\beta - \beta(h_q)) \right], \end{aligned}$$

where $\mathcal{L}_q(\cdot; h_q)$ denotes the perturbed auxiliary criterion for equation q . Both terms on the right-hand side are nonnegative. When A^c is nonempty, the strict inequality $\|z_{A^c}(h_q)\|_{\infty} < 1$ implies that equality can hold only if

$$\beta_{A^c} = 0.$$

When A^c is empty, this restriction holds trivially. Conditional on this restriction, equality in the first term requires

$$\tilde{X}_A^{(k)} (\beta_A - \beta_A(h_q)) = 0.$$

Since Q_A is positive definite, $\tilde{X}_A^{(k)}$ has full column rank. Hence,

$$\beta_A = \beta_A(h_q).$$

The perturbed minimizer is therefore unique and has the same nonzero coefficient index set A .

Now suppose that $\mathcal{A}_q^{(k)}$ is empty. In this case, the original minimizer is the zero vector. Define the perturbed score

$$g_q(h_q) = \frac{1}{T_{\mathcal{L}}} (\tilde{X}^{(k)})' (\tilde{\mathbf{u}}_q^{(k)} + M h_q).$$

Assumption 2 gives

$$\|g_q(0)\|_{\infty} \leq \lambda - \delta_q$$

for some $\delta_q > 0$. The mapping $h_q \mapsto g_q(h_q)$ is continuous. There therefore exists $\epsilon_q^0 > 0$ such that

$$\|h_q\|_2 < \epsilon_q^0$$

implies

$$\|g_q(h_q)\|_{\infty} < \lambda.$$

For any nonzero coefficient vector β , the difference between the perturbed criterion at β and at zero is

$$\begin{aligned} \mathcal{L}_q(\beta; h_q) - \mathcal{L}_q(0; h_q) &= \frac{1}{2T_{\mathcal{L}}} \left\| \tilde{X}^{(k)} \beta \right\|_2^2 - g_q(h_q)' \beta + \lambda \|\beta\|_1 \\ &\geq \frac{1}{2T_{\mathcal{L}}} \left\| \tilde{X}^{(k)} \beta \right\|_2^2 + (\lambda - \|g_q(h_q)\|_{\infty}) \|\beta\|_1 \\ &> 0. \end{aligned}$$

The zero vector therefore remains the unique minimizer, and its nonzero coefficient index set remains empty.

For each equation q , let ϵ_q equal ϵ_q^A when $\mathcal{A}_q^{(k)}$ is nonempty and ϵ_q^0 when $\mathcal{A}_q^{(k)}$ is empty. Each ϵ_q is strictly positive. Since the current system contains finitely many equations, define

$$\epsilon_k = \min_{q \in \mathcal{Q}} \epsilon_q > 0.$$

If $\|h_q\|_2 < \epsilon_k$ for every equation q , each nonzero coefficient index set $\mathcal{A}_q^{(k)}$ remains unchanged. The coefficient blocks in (13) are therefore unchanged in their zero and nonzero pattern. It follows from (14) that $\widetilde{\mathcal{S}}^{(k)}(\lambda)$ is unchanged, and (15) then implies that $\mathcal{S}^{(k+1)}$ is unchanged. $\square$

Proof of Corollary 1. Let $\{\mathcal{S}^{(k)}\}_{k=0}^{\bar{k}}$ denote the original construction path, and let $\{\mathcal{S}_h^{(k)}\}$ denote the path generated by the perturbed fitted composite disturbances. Both procedures begin from the same core index set. Hence,

$$\mathcal{S}_h^{(0)} = \mathcal{S}^{(0)}.$$

Suppose for some $k \leq \bar{k}$ that

$$\mathcal{S}_h^{(k)} = \mathcal{S}^{(k)}.$$

The current system and the set of remaining candidates are then the same along the two paths. Because

$$\|h_q^{(k)}\|_2 < \epsilon_k \quad \text{for every } q \in \mathcal{Q},$$

Proposition 5 implies

$$\widetilde{\mathcal{S}}_h^{(k)}(\lambda) = \widetilde{\mathcal{S}}^{(k)}(\lambda).$$

Applying the update rule (15) therefore gives

$$\mathcal{S}_h^{(k+1)} = \mathcal{S}^{(k+1)}$$

whenever the original procedure continues beyond iteration k . By induction, the perturbed and original index sets coincide at every iteration through $\bar{k}$.

It remains to show that the terminal iteration is also unchanged. For every $k < \bar{k}$, the definition of $\bar{k}$ gives

$$\widetilde{\mathcal{S}}^{(k)}(\lambda) \neq \emptyset.$$

Since the perturbed selected set is identical, the perturbed procedure cannot terminate before $\bar{k}$. At iteration $\bar{k}$,

$$\widetilde{\mathcal{S}}^{(\bar{k})}(\lambda) = \emptyset.$$

Proposition 5 gives

$$\widetilde{\mathcal{S}}_h^{(\bar{k})}(\lambda) = \emptyset,$$

Thus, the perturbed procedure also terminates at $\bar{k}$. Proposition 1 ensures that $\bar{k}$ is finite. The complete construction path and its terminal index set are therefore unchanged. Hence,

$$y_{t,h}^\dagger(\lambda) = y_t^\dagger(\lambda).$$

$\square$

Proof of Lemma 2. Fix the loss g . By Assumption 3(iii), for every $\lambda \in \Lambda$ the posterior mean loss

$$\overline{\mathcal{L}}_g^{(T)}(\lambda) = \frac{1}{S} \sum_{s=1}^S \mathcal{L}_g^{(s,T)}(\lambda)$$

is a finite average of finite numbers and hence a finite real number. The image $\{\overline{\mathcal{L}}_g^{(T)}(\lambda) : \lambda \in \Lambda\}$ is a finite subset of $\mathbb{R}$ and therefore attains a minimum m_g . The minimizer set

$$M_g = \left\{ \lambda \in \Lambda : \overline{\mathcal{L}}_g^{(T)}(\lambda) = m_g \right\}$$

is nonempty and finite. By the strict total order $\prec$ in Assumption 3(i), M_g has a unique $\prec$ -least element. Set

$$\lambda_g^* = \min_{\prec} M_g,$$

as in (20) applied under loss g . This proves existence and uniqueness of the reference model complexity parameter. Assumption 3(ii) then ensures that the associated reference terminal system is also unique.

Given λ_g^* , for every $\lambda \in \Lambda \setminus \{\lambda_g^*\}$ the paired differences in (21) are

$$\Delta_g^{(s,T)}(\lambda) = \mathcal{L}_g^{(s,T)}(\lambda) - \mathcal{L}_g^{(s,T)}(\lambda_g^*), \quad s = 1, \dots, S.$$

The common index s labels a pair consisting of one posterior draw from each terminal system. By Assumption 3(iii), both loss values in each pair are finite. Hence, $\Delta_g^{(s,T)}(\lambda)$ is finite for every $\lambda \in \Lambda \setminus \{\lambda_g^*\}$ and every $s = 1, \dots, S$.

Because $\alpha \in (0, 1)$,

$$\frac{1-\alpha}{2} \in \left(0, \frac{1}{2}\right), \quad \frac{1+\alpha}{2} \in \left(\frac{1}{2}, 1\right).$$

By Assumption 3(iv), the endpoints $q_{(1-\alpha)/2}^g(\lambda)$ and $q_{(1+\alpha)/2}^g(\lambda)$ are uniquely determined for every $\lambda \in \Lambda \setminus \{\lambda_g^*\}$. Consequently, each of the conditions

$$q_{(1+\alpha)/2}^g(\lambda) < 0 \quad \text{and} \quad q_{(1-\alpha)/2}^g(\lambda) \leq 0 \leq q_{(1+\alpha)/2}^g(\lambda)$$

is either true or false for every such λ , with no indeterminacy. The sets

$$\Lambda_\alpha^{g,-} = \left\{ \lambda \in \Lambda \setminus \{\lambda_g^*\} : q_{(1+\alpha)/2}^g(\lambda) < 0 \right\}$$

and

$$\Lambda_\alpha^{g,0} = \{\lambda_g^*\} \cup \left\{ \lambda \in \Lambda \setminus \{\lambda_g^*\} : q_{(1-\alpha)/2}^g(\lambda) \leq 0 \leq q_{(1+\alpha)/2}^g(\lambda) \right\}$$

are therefore uniquely determined subsets of Λ .

The explicit inclusion of λ_g^* in $\Lambda_\alpha^{g,0}$ reflects that the OOS forecast loss difference between the reference system and itself is identically zero. No empirical quantile comparison is required for the reference system. Hence, $\lambda_g^* \in \Lambda_\alpha^{g,0}$ and $\Lambda_\alpha^{g,0} \neq \emptyset$.

Finally, by (22) applied under loss g ,

$$\Lambda_\alpha^g = \begin{cases} \Lambda_\alpha^{g,-}, & \Lambda_\alpha^{g,-} \neq \emptyset, \\ \Lambda_\alpha^{g,0}, & \Lambda_\alpha^{g,-} = \emptyset. \end{cases}$$

In the first case, $\Lambda_\alpha^g = \Lambda_\alpha^{g,-}$ is nonempty by the case condition. In the second case, $\Lambda_\alpha^g = \Lambda_\alpha^{g,0}$ is nonempty because it contains λ_g^* . Both branches are uniquely determined by the sets constructed above. Therefore, Λ_α^g exists, is unique, and is nonempty. $\square$

Proof of Proposition 6. Apply Lemma 2 first with $g = \text{SE}$ and then with $g = \text{AE}$. The retained sets $\Lambda_\alpha^{\text{SE}}$ and $\Lambda_\alpha^{\text{AE}}$ therefore exist, are unique, and are nonempty. Their union

$$\Lambda_\alpha^\cup = \Lambda_\alpha^{\text{SE}} \cup \Lambda_\alpha^{\text{AE}}$$

is consequently a uniquely determined subset of Λ . It is nonempty because it contains the nonempty sets $\Lambda_\alpha^{\text{SE}}$ and $\Lambda_\alpha^{\text{AE}}$. It is finite because it is a subset of the finite grid Λ . Hence, the combined retained set $\Lambda_\alpha^\cup$ exists, is unique, is nonempty, and is finite.

By Assumption 3(ii), the terminal dimension $n^\dagger(\lambda)$ is well defined for every $\lambda \in \Lambda$. It is therefore well defined for every $\lambda \in \Lambda_\alpha^\cup$. Because $\Lambda_\alpha^\cup$ is finite and nonempty, the image

$$\{n^\dagger(\lambda) : \lambda \in \Lambda_\alpha^\cup\}$$

is a finite nonempty set of nonnegative integers. Every finite nonempty set of integers has a unique largest value. Thus,

$$n_{\alpha,\cup}^{\max} = \max_{\lambda \in \Lambda_\alpha^\cup} n^\dagger(\lambda)$$

exists and is unique.

Given the uniquely determined value $n_{\alpha,\cup}^{\max}$, define

$$\Lambda_{\alpha,\cup}^{\max} = \{\lambda \in \Lambda_\alpha^\cup : n^\dagger(\lambda) = n_{\alpha,\cup}^{\max}\}.$$

This set is uniquely determined because both $\Lambda_\alpha^\cup$ and the function $\lambda \mapsto n^\dagger(\lambda)$ are uniquely determined. It is nonempty because the maximum $n_{\alpha,\cup}^{\max}$ is attained by at least one member of the finite nonempty set $\Lambda_\alpha^\cup$. It is finite because it is a subset of the finite set $\Lambda_\alpha^\cup$. Hence, $\Lambda_{\alpha,\cup}^{\max}$ exists, is unique, is nonempty, and is finite.

The strict total order $\prec$ on Λ in Assumption 3(i) restricts to a strict total order on the finite nonempty subset $\Lambda_{\alpha,\cup}^{\max}$. Every finite nonempty set endowed with a strict total order has a unique least element. Therefore,

$$\hat{\lambda}_\alpha = \min_{\prec} \Lambda_{\alpha,\cup}^{\max}$$

exists and is unique.

Finally, Assumption 3(ii) states that the terminal variable vector $y_t^\dagger(\lambda)$ is unique for every $\lambda \in \Lambda$. Since $\hat{\lambda}_\alpha \in \Lambda_{\alpha,\cup}^{\max} \subseteq \Lambda$ exists and is unique, the terminal system $y_t^\dagger(\hat{\lambda}_\alpha)$ also exists and is unique. Thus, the selected terminal system is uniquely determined by the data and the prespecified procedure. $\square$

APPENDIX S2. PROXY IDENTIFICATION AND THE JOINT BAYESIAN METHOD

This appendix proves the results that Section VII defers to it. It proves Proposition 7 and Theorem 1, derives the two-stage coefficients in (28) and proves Proposition 8, shows the rotational invariance of the remaining equations in posterior computation, and states the efficiency comparison under correct parametric specification.

Proof of Proposition 7. By (24),

$$v_t = A_0^{-1'} \varepsilon_t.$$

Therefore,

$$G \equiv \mathbb{E}(v_t m_t') = A_0^{-1'} \mathbb{E}(\varepsilon_t m_t').$$

Assumption 4 restricts the rows of $\mathbb{E}(\varepsilon_t m_t')$. Its j th row is

$$\mathbb{E}(\varepsilon_{jt} m_t') = (\mathbb{E}(m_t \varepsilon_{jt}))'.$$

By (26), the first row equals γ_m' , while every remaining row equals $0_{1 \times n_m}$. Stacking these rows gives

$$\mathbb{E}(\varepsilon_t m_t') = \begin{pmatrix} \gamma_m' \\ 0_{1 \times n_m} \\ \vdots \\ 0_{1 \times n_m} \end{pmatrix} = e_1 \gamma_m'.$$

This result holds for every $n_m \geq 1$, because Assumption 4 restricts which shocks may be correlated with the instruments, not how many instruments there are.

Using the definition of s in (25),

$$G = A_0^{-1'} e_1 \gamma_m' = s \gamma_m'.$$

Recall that

$$G' = [g_1 \ \cdots \ g_n].$$

The i th row of $G = s \gamma_m'$ is $s_i \gamma_m'$. Hence,

$$g_i = \gamma_m s_i, \quad i = 1, \dots, n.$$

Let p be an admissible anchor. By definition,

$$g_p \neq 0_{n_m \times 1}.$$

Since $g_p = \gamma_m s_p$ and $\gamma_m \neq 0_{n_m \times 1}$ under Assumption 4, admissibility implies

$$s_p \neq 0.$$

Let

$$W = \mathbb{E}(m_t m_t')^{-1}.$$

Assumption 4 states that $\mathbb{E}(m_t m_t')$ is positive definite. Hence, W is positive definite and

$$\gamma_m' W \gamma_m > 0.$$

By (28), the two-stage coefficient for variable i using anchor p is

$$\beta_{i|p} = \frac{g'_p W g_i}{g'_p W g_p}.$$

Substituting $g_i = \gamma_m s_i$ and $g_p = \gamma_m s_p$ gives

$$\beta_{i|p} = \frac{s_p s_i \gamma'_m W \gamma_m}{s_p^2 \gamma'_m W \gamma_m} = \frac{s_i}{s_p}.$$

In particular,

$$\beta_{p|p} = 1.$$

Collecting these coefficients gives

$$d_p = \begin{pmatrix} \beta_{1|p} \\ \vdots \\ \beta_{n|p} \end{pmatrix} = \frac{1}{s_p} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \frac{s}{s_p}.$$

Let r be another admissible anchor. The same argument gives

$$d_r = \frac{s}{s_r}.$$

Because $s_p \neq 0$ and $s_r \neq 0$,

$$d_r = \frac{s}{s_r} = \frac{s_p}{s_r} \frac{s}{s_p} = \frac{s_p}{s_r} d_p.$$

Thus, ratio vectors obtained from different admissible anchors are proportional. The anchor determines which component of the ratio vector is normalized to one.

It remains to show that the scale normalization removes this anchor-specific representation. By (24),

$$\Sigma = A_0^{-1'} A_0^{-1},$$

and hence

$$\Sigma^{-1} = A_0 A_0'.$$

Using $s = A_0^{-1'} e_1$ from (25),

$$s' \Sigma^{-1} s = e_1' A_0^{-1} A_0 A_0' A_0^{-1'} e_1 = e_1' e_1 = 1.$$

Since $d_p = s/s_p$,

$$d_p' \Sigma^{-1} d_p = \frac{1}{s_p^2} s' \Sigma^{-1} s = \frac{1}{s_p^2}.$$

Therefore,

$$\sqrt{d_p' \Sigma^{-1} d_p} = \frac{1}{|s_p|}.$$

The unit-variance normalization in (29) then gives

$$\frac{d_p}{\sqrt{d_p' \Sigma^{-1} d_p}} = \frac{s/s_p}{1/|s_p|} = \frac{|s_p|}{s_p} s = \text{sgn}(s_p) s.$$

Thus, every admissible anchor recovers the same impact vector up to sign. Applying the same economic sign normalization selects the common orientation s for every anchor.

The matrices Ψ_h in (37) are generated by the reduced-form matrices $B_\ell = A_\ell A_0^{-1}$ in (36), which are identified from the reduced form without reference to any anchor. The impulse responses therefore satisfy

$$\text{IRF}_1(h) = s' \Psi_h.$$

Because every admissible anchor recovers the same sign-normalized impact vector s , every admissible anchor produces the same impulse responses at every horizon. Moreover, $a_1 = \Sigma^{-1}s$ by (25), so every admissible anchor also recovers the same first equation. $\square$

Proof of Theorem 1. Transposing (24) gives

$$v_t = A_0^{-1'} \varepsilon_t.$$

The maintained exclusion restrictions imply

$$\mathbb{E}(\varepsilon_t m'_t) = e_1 \gamma'_m.$$

Therefore,

$$G = \mathbb{E}(v_t m'_t) = A_0^{-1'} \mathbb{E}(\varepsilon_t m'_t) = A_0^{-1'} e_1 \gamma'_m = s \gamma'_m.$$

Since

$$\Sigma^{-1} = A_0 A'_0 \quad \text{and} \quad s = A_0^{-1'} e_1,$$

we have

$$\Sigma^{-1} s = A_0 A'_0 A_0^{-1'} e_1 = A_0 e_1 = a_1.$$

Hence,

$$\Sigma^{-1} G = a_1 \gamma'_m.$$

Because A_0 is invertible, both a_1 and s are nonzero. It follows that

$$\gamma_m \neq 0_{n_m \times 1} \iff G \neq 0_{n \times n_m} \iff \text{rank}(G) = 1.$$

Since Σ^{-1} is invertible,

$$\text{rank}(\Sigma^{-1} G) = \text{rank}(G).$$

Thus, the relevance condition is equivalent to the two rank conditions stated in the theorem.

Suppose first that $\gamma_m \neq 0_{n_m \times 1}$. Let $q \in \mathbb{R}^{n_m}$ satisfy $Gq \neq 0_{n \times 1}$. From (34),

$$Gq = (\gamma'_m q) s, \quad \Sigma^{-1} Gq = (\gamma'_m q) a_1.$$

Since $s \neq 0_{n \times 1}$, the condition $Gq \neq 0_{n \times 1}$ implies $\gamma'_m q \neq 0$. The unit-variance normalization of the first shock gives

$$a'_1 \Sigma a_1 = e'_1 A'_0 A_0^{-1'} A_0^{-1} A_0 e_1 = e'_1 e_1 = 1.$$

Therefore,

$$(\Sigma^{-1} Gq)' \Sigma (\Sigma^{-1} Gq) = (\gamma'_m q)^2,$$

and

$$\frac{\Sigma^{-1}Gq}{\sqrt{(\Sigma^{-1}Gq)' \Sigma (\Sigma^{-1}Gq)}} = \text{sgn}(\gamma'_m q) a_1.$$

This proves the expression for a_1 in (35). Moreover,

$$\Sigma a_1 = A_0^{-1'} A_0^{-1} A_0 e_1 = A_0^{-1'} e_1 = s,$$

which implies

$$s' \Sigma^{-1} s = a_1' \Sigma a_1 = 1.$$

Using $Gq = (\gamma'_m q)s$ gives

$$(Gq)' \Sigma^{-1} (Gq) = (\gamma'_m q)^2,$$

and hence

$$\frac{Gq}{\sqrt{(Gq)' \Sigma^{-1} (Gq)}} = \text{sgn}(\gamma'_m q) s.$$

This proves the expression for s in (35).

Let

$$B_\ell = A_\ell A_0^{-1}, \quad \ell = 1, \dots, \mathcal{L},$$

and let

$$\tilde{c}' = c' A_0^{-1}.$$

Since $a_1 = A_0 e_1$,

$$a_{\ell,1} = A_\ell e_1 = B_\ell a_1, \quad \ell = 1, \dots, \mathcal{L},$$

and

$$c_1 = c' e_1 = \tilde{c}' a_1.$$

Thus, identification of a_1 identifies all contemporaneous, lagged, and constant coefficients in the first equation under the same sign normalization.

The factorization $G = s\gamma'_m$ and the normalization $s' \Sigma^{-1} s = 1$ imply

$$\gamma'_m = s' \Sigma^{-1} G.$$

Hence,

$$\delta_m = \Omega_m^{-1} \gamma_m$$

is also identified under the common sign normalization. The impulse responses to the first shock satisfy

$$\text{IRF}_1(h) = s' \Psi_h, \quad h \geq 0.$$

Since s and the reduced-form dynamic coefficients determining Ψ_h are identified, the impulse responses to the first shock are identified at every horizon. This proves sufficiency.

It remains to prove necessity. Suppose that

$$\gamma_m = 0_{n_m \times 1}.$$

Then

$$\mathbb{E}(\varepsilon_t m'_t) = 0_{n \times n_m} \quad \text{and} \quad G = 0_{n \times n_m}.$$

Because $n \geq 2$, there exists an orthogonal $n \times n$ matrix P such that $Pe_1 \neq e_1$ and $Pe_1 \neq -e_1$. Define

$$\begin{aligned} A_\ell^* &= A_\ell P, & \ell &= 0, \dots, \mathcal{L}, \\ c^* &= P'c, & \varepsilon_t^* &= P'\varepsilon_t. \end{aligned}$$

Orthogonality gives

$$\mathbb{E}(\varepsilon_t^* \varepsilon_t^{*'}) = I_n,$$

and

$$\mathbb{E}(\varepsilon_t^* m_t') = P' \mathbb{E}(\varepsilon_t m_t') = 0_{n \times n_m}.$$

Thus, the transformed representation satisfies the same proxy restrictions. Its reduced-form coefficients are unchanged because

$$A_\ell^* (A_0^*)^{-1} = A_\ell P P' A_0^{-1} = A_\ell A_0^{-1}, \quad \ell = 1, \dots, \mathcal{L},$$

and

$$c^{*'} (A_0^*)^{-1} = c' P P' A_0^{-1} = c' A_0^{-1}.$$

The reduced-form innovation is also unchanged:

$$\varepsilon_t^{*'} (A_0^*)^{-1} = \varepsilon_t' P P' A_0^{-1} = \varepsilon_t' A_0^{-1}.$$

Hence, the original and transformed systems imply the same distribution for the observed variables and instruments. Their first contemporaneous coefficient vectors nevertheless differ:

$$a_1^* = A_0^* e_1 = A_0 P e_1.$$

Because A_0 is invertible and $Pe_1 \neq \pm e_1$,

$$a_1^* \neq \pm a_1.$$

Thus, when $\gamma_m = 0_{n_m \times 1}$, two observationally equivalent parameterizations satisfy the same proxy restrictions but have first equations that differ by more than a common sign. The first equation is therefore not identified. This proves necessity and establishes the if-and-only-if result.

Let Q be any orthogonal $(n-1) \times (n-1)$ matrix and define

$$H = \text{diag}(1, Q).$$

Postmultiplying A_ℓ by H for $\ell = 0, \dots, \mathcal{L}$, replacing c by $H'c$, and replacing ε_t by $H'\varepsilon_t$ leaves the reduced-form unchanged. Since $He_1 = e_1$, this transformation also leaves the first equation, the first shock, the proxy restrictions, and the impulse responses to the first shock unchanged, while it generally changes the remaining equations. The remaining equations are therefore not identified without additional restrictions and are not needed for the identified impulse responses.

Last, Proposition S1 shows that orthogonal transformations of columns 2, $\dots$, n leave the joint covariance restriction, the first equation, and the impulse responses to the first shock

unchanged. Thus, the remaining equations are not identified and are not needed for these impulse responses. $\square$

The two-stage coefficients. The main text states the two-stage coefficient in (28) without derivation. We derive it here from population moments, using G , g_i , and $W = \Omega_m^{-1}$ as defined in Section VII.1. Fix an admissible anchor p . The linear projection of the anchor innovation on the instrument vector is

$$v_{p,t} = m_t' \pi_p + \zeta_{p,t}, \quad \pi_p = W g_p, \quad \mathbb{E}(m_t \zeta_{p,t}) = 0_{n_m \times 1},$$

where $\zeta_{p,t}$ is the projection residual. The fitted value is $\tilde{v}_{p,t} = m_t' W g_p$. The population two-stage least squares coefficient from regressing $v_{i,t}$ on $v_{p,t}$ using m_t as the instrument vector is the ratio of the covariance of $v_{i,t}$ with the fitted value to the covariance of $v_{p,t}$ with the fitted value. Since

$$\mathbb{E}(\tilde{v}_{p,t} v_{i,t}) = g_p' W g_i \quad \text{and} \quad \mathbb{E}(\tilde{v}_{p,t} v_{p,t}) = g_p' W g_p,$$

we obtain

$$\beta_{i|p} = \frac{\mathbb{E}(\tilde{v}_{p,t} v_{i,t})}{\mathbb{E}(\tilde{v}_{p,t} v_{p,t})} = \frac{g_p' W g_i}{g_p' W g_p},$$

which is (28). The denominator is positive because W is positive definite and $g_p \neq 0_{n_m \times 1}$. Under Assumption 4, an anchor is admissible if and only if $s_p \neq 0$, because (34) gives $g_p = \gamma_m s_p$ and $\gamma_m \neq 0_{n_m \times 1}$.

Proof of Proposition 8. By (34), $G = s \gamma_m'$, which gives $g_i = \gamma_m s_i$ for every i . For any admissible anchor p ,

$$\beta_{i|p} = \frac{g_p' W g_i}{g_p' W g_p} = \frac{s_p s_i \gamma_m' W \gamma_m}{s_p^2 \gamma_m' W \gamma_m} = \frac{s_i}{s_p},$$

where $\gamma_m' W \gamma_m > 0$ because W is positive definite and $\gamma_m \neq 0_{n_m \times 1}$. Collecting these coefficients gives $d_p = s/s_p$.

The unit-variance normalization gives $s' \Sigma^{-1} s = 1$, as established in the proof of Theorem 1. Therefore,

$$d_p' \Sigma^{-1} d_p = \frac{s' \Sigma^{-1} s}{s_p^2} = \frac{1}{s_p^2}, \quad \frac{d_p}{\sqrt{d_p' \Sigma^{-1} d_p}} = \text{sgn}(s_p) s.$$

The scale normalization in (29) therefore returns s up to sign, and the economic sign normalization selects the same s for every admissible anchor.

For the joint procedure, Theorem 1 identifies s directly from G and Σ . For any $q \in \mathbb{R}^{n_m}$ such that $Gq \neq 0_{n \times 1}$, the identity $Gq = (\gamma_m' q)s$ and the normalization $s' \Sigma^{-1} s = 1$ reduce the second expression in (35) to $\pm s$. The same economic sign normalization selects the same s . Both procedures therefore identify the same impact vector.

The two procedures also identify the same first equation. Since $a_1 = \Sigma^{-1}s$, they identify the same contemporaneous coefficient vector. The reduced-form matrices $B_\ell = A_\ell A_0^{-1}$ are common to both procedures, and hence

$$a_{\ell,1} = A_\ell e_1 = B_\ell a_1, \quad \ell = 1, \dots, \mathcal{L},$$

is the same under both. Likewise,

$$c_1 = c'e_1 = c'A_0^{-1}a_1,$$

and $c'A_0^{-1}$ is the reduced-form constant common to both procedures. Thus, every coefficient in the first equation agrees after the common sign normalization.

Finally, the moving-average matrices Ψ_h in (37) are functions of the common reduced-form matrices. Equation (38) therefore gives

$$\text{IRF}_1(h) = s'\Psi_h$$

for both procedures at every horizon $h \geq 0$. □

Two features of this equivalence deserve emphasis. First, the anchor enters the two-stage construction even though the identified object does not depend on it. The fitted value $\tilde{v}_{p,t} = s_p m_t' W \gamma_m$ and the ratio vector $d_p = s/s_p$ vary with p , and a second admissible anchor r gives $d_r = (s_p/s_r)d_p$. Under the rank-one restriction, these representations are proportional, and the common scale and sign normalization maps each of them to the same s . Second, Proposition 7 characterizes the role of this rank-one restriction. When it fails, ratio vectors based on different anchors need not be proportional and, in general, no common normalization reconciles them. The equivalence in Proposition 8 is therefore a consequence of the maintained proxy restrictions, not a generic property of the two-stage procedure.

The same distinction has a finite-sample counterpart. Let $\hat{G}$ denote the sample counterpart of G , formed from the estimated reduced-form innovations, and let $\hat{g}_i$ denote the transpose of its i th row. With one instrument, each $\hat{g}_i$ is a scalar and

$$\hat{\beta}_{i|p} = \frac{\hat{g}_i}{\hat{g}_p}$$

whenever $\hat{g}_p \neq 0$, so changing the anchor only rescales the same estimated direction. With more than one instrument, sampling variation generally makes $\hat{G}$ have rank greater than one even when G has rank one. The vectors $\hat{g}_1, \dots, \hat{g}_n$ then need not share a common direction, and

$$\hat{\beta}_{i|p} = \frac{\hat{g}_p' \hat{W} \hat{g}_i}{\hat{g}_p' \hat{W} \hat{g}_p}$$

depends on the anchor through $\hat{g}_p$. Different anchors therefore generally produce estimated directions that are not proportional. This finite-sample anchor dependence can arise even when the population proxy restrictions are correct, and it is distinct from a population failure of the rank-one restriction. The joint method avoids anchor dependence because it neither

selects a residual anchor nor divides by an estimated anchor loading. It does not remove the need for valid and relevant instruments or for the rank-one restriction to hold in population.

The following propositions establish two implications of the joint formulation beyond identification. Proposition S1 shows that rotations of the remaining equations leave the identified equation, shock, and impulse responses unchanged. Proposition S2 shows that joint estimation attains the asymptotic efficiency bound and weakly dominates two-stage estimation of the same object under the stated regularity conditions.

Proposition S1. *Suppose that $n \geq 2$, the joint system satisfies (30), and $\gamma_m \neq 0_{n_m \times 1}$. Let Q be any orthogonal $(n-1) \times (n-1)$ matrix and define*

$$H = \text{diag}(1, Q).$$

For $\ell = 0, \dots, \mathcal{L}$, define

$$A_\ell^* = A_\ell H, \quad c^* = H'c, \quad \varepsilon_t^* = H'\varepsilon_t.$$

Then the transformed coefficients and shocks imply the same reduced form and satisfy the same joint covariance restriction as the original joint system. The transformation leaves the first equation, the first shock, and the impulse responses to that shock unchanged at every horizon. Consequently, the equations associated with columns $2, \dots, n$ are not identified without additional restrictions, and their rotational indeterminacy does not affect identification of the first equation or its impulse responses.

Proof. Theorem 4 of Zha (1999) establishes the block-rotation invariance underlying this result. The first equation forms one block, while the equations associated with columns $2, \dots, n$ form the transformed block. Because the SVAR is written in row-vector form, the premultiplication in Zha (1999) appears here as postmultiplication of the coefficient matrices by H . We verify the result directly and then impose the external-instrument covariance restriction.

The matrix H is orthogonal and leaves the first coordinate unchanged:

$$H'H = HH' = I_n, \quad He_1 = H'e_1 = e_1.$$

Postmultiplying (1) by H gives the transformed structural representation because

$$A_\ell^* = A_\ell H, \quad c^{*'} = c'H, \quad \varepsilon_t^{*'} = \varepsilon_t'H.$$

For each $\ell = 1, \dots, \mathcal{L}$,

$$A_\ell^*(A_0^*)^{-1} = A_\ell H(A_0 H)^{-1} = A_\ell H H' A_0^{-1} = A_\ell A_0^{-1}.$$

The reduced-form constant is also unchanged:

$$c^{*'}(A_0^*)^{-1} = c'H H' A_0^{-1} = c'A_0^{-1}.$$

Under the transformed representation, the reduced-form innovation satisfies

$$v_t^{*'} = \varepsilon_t^{*'}(A_0^*)^{-1} = \varepsilon_t' H H' A_0^{-1} = \varepsilon_t' A_0^{-1} = v_t'.$$

Thus, the observed-data reduced form is unchanged.

We next verify the joint covariance restriction. Since $\varepsilon_t^* = H' \varepsilon_t$,

$$\mathbb{E}(\varepsilon_t^* \varepsilon_t^{*'}) = H' \mathbb{E}(\varepsilon_t \varepsilon_t') H = I_n.$$

Equation (30) gives $\mathbb{E}(\varepsilon_t m_t') = e_1 \gamma_m'$, and hence

$$\mathbb{E}(\varepsilon_t^* m_t') = H' \mathbb{E}(\varepsilon_t m_t') = H' e_1 \gamma_m' = e_1 \gamma_m'.$$

Because m_t is not transformed, its covariance remains Ω_m . Therefore,

$$\mathbb{E} \left(\begin{bmatrix} \varepsilon_t^* \\ m_t \end{bmatrix} \begin{bmatrix} \varepsilon_t^{*'} & m_t' \end{bmatrix} \right) = \begin{bmatrix} I_n & e_1 \gamma_m' \\ \gamma_m e_1' & \Omega_m \end{bmatrix}.$$

Thus, the transformed representation satisfies the same joint covariance restriction.

The first column of every coefficient matrix is unchanged. For $\ell = 0, \dots, \mathcal{L}$,

$$A_\ell^* e_1 = A_\ell H e_1 = A_\ell e_1.$$

In particular,

$$a_1^* = a_1, \quad a_{\ell,1}^* = a_{\ell,1}, \quad \ell = 1, \dots, \mathcal{L}.$$

The constant in the first equation is also unchanged:

$$c_1^* = e_1' c^* = e_1' H' c = e_1' c = c_1.$$

The first shock is unchanged as well:

$$\varepsilon_{1t}^* = e_1' \varepsilon_t^* = e_1' H' \varepsilon_t = \varepsilon_{1t}.$$

Because γ_m and Ω_m are unchanged, $\delta_m = \Omega_m^{-1} \gamma_m$ is unchanged. Hence the projection residual

$$w_{1t} = \varepsilon_{1t} - m_t' \delta_m$$

is also unchanged. The augmented first equation (33) is therefore identical under the original and transformed representations. Moreover, (30) implies

$$\mathbb{E}(w_{1t} \varepsilon_{-1,t}') = 0_{1 \times (n-1)},$$

where $\varepsilon_{-1,t}$ collects shocks $2, \dots, n$. Since $\varepsilon_{-1,t}^* = Q' \varepsilon_{-1,t}$,

$$\mathbb{E}(w_{1t} \varepsilon_{-1,t}^{*'}) = \mathbb{E}(w_{1t} \varepsilon_{-1,t}') Q = 0_{1 \times (n-1)}.$$

Thus, the orthogonality conditions in the joint system are preserved without invoking any assumption beyond (30).

It remains to verify the impulse responses. Let $s' = e_1' A_0^{-1}$. Under the transformed representation,

$$s^{*'} = e_1' (A_0^*)^{-1} = e_1' H' A_0^{-1} = e_1' A_0^{-1} = s'.$$

The reduced-form coefficient matrices are unchanged, so the moving-average matrices satisfy $\Psi_h^* = \Psi_h$ for every $h \geq 0$. Therefore,

$$\text{IRF}_1^*(h) = s'^* \Psi_h^* = s' \Psi_h = \text{IRF}_1(h) \quad \text{for every } h \geq 0.$$

This is the specialization of the block invariance result in Theorem 4 of [Zha \(1999\)](#) to the first shock and the block of remaining equations. The external-instrument covariance restriction is preserved because the transformation fixes the first coordinate. The remaining equations can vary across observationally equivalent representations. Under the maintained relevance condition, Theorem 1 identifies the first equation up to sign, and the transformation above leaves that equation, the first shock, and its impulse responses unchanged. $\square$

Proposition S2. *Suppose that the Gaussian joint proxy SVAR is correctly specified and admits a finite-dimensional, locally identified regular parameterization ϑ in a neighborhood of an interior true value ϑ_0 . Assume that the likelihood is locally asymptotically normal with nonsingular Fisher information $\mathcal{I}(\vartheta_0)$, that the posterior satisfies the Bernstein–von Mises conclusion ([van der Vaart, 1998](#), Theorem 10.1) centered at the maximum-likelihood estimator $\hat{\vartheta}_{\text{ML}}$, and that the prior density is positive and continuously differentiable in a neighborhood of ϑ_0 . Assume further that there exists a sequence $\delta_T \downarrow 0$ with $\sqrt{T}\delta_T \rightarrow \infty$ such that, with probability approaching one, an interior posterior mode $\hat{\vartheta}_{\text{PM}}$ and $\hat{\vartheta}_{\text{ML}}$ lie in $B(\vartheta_0, \delta_T)$, that $T^{-1}\nabla^2 \ell_T(\vartheta)$ converges uniformly to $-\mathcal{I}(\vartheta_0)$ on this neighborhood, and that the log-prior score is bounded there. Let $\psi(\vartheta)$ be any continuously differentiable identified object common to the joint and two-stage methods, including a_1 , s , or a finite collection of impulse responses to the first shock. Then the joint posterior for $\sqrt{T}\{\psi(\vartheta) - \psi(\hat{\vartheta}_{\text{ML}})\}$ converges to*

$$N(0, V_{\text{eff}}), \quad V_{\text{eff}} = D\psi(\vartheta_0)\mathcal{I}(\vartheta_0)^{-1}D\psi(\vartheta_0)'.$$

Under the mode-localization and uniform-Hessian conditions,

$$\hat{\vartheta}_{\text{PM}} - \hat{\vartheta}_{\text{ML}} = O_p(T^{-1}) = o_p(T^{-1/2}).$$

If a two-stage estimator $\hat{\psi}_{2\text{S}}$ is regular and asymptotically linear for the same object under the same joint model and has asymptotic covariance $V_{2\text{S}}$, then

$$V_{2\text{S}} - V_{\text{eff}}$$

is positive semidefinite. Equality holds if and only if the influence function of $\hat{\psi}_{2\text{S}}$ equals the efficient influence function almost surely.

Proof. Let $b(\vartheta) = \nabla \log \pi(\vartheta)$ denote the log-prior score. The first-order conditions for the posterior mode and the maximum-likelihood estimator imply

$$0 = \nabla \ell_T(\hat{\vartheta}_{\text{PM}}) + b(\hat{\vartheta}_{\text{PM}}) = \nabla^2 \ell_T(\tilde{\vartheta})(\hat{\vartheta}_{\text{PM}} - \hat{\vartheta}_{\text{ML}}) + b(\hat{\vartheta}_{\text{PM}})$$

for some $\tilde{\vartheta}$ on the line segment joining $\hat{\vartheta}_{\text{PM}}$ and $\hat{\vartheta}_{\text{ML}}$. The uniform Hessian limit, nonsingularity of $\mathcal{I}(\vartheta_0)$, and bounded log-prior score therefore yield

$$\hat{\vartheta}_{\text{PM}} - \hat{\vartheta}_{\text{ML}} = O_p(T^{-1}) = o_p(T^{-1/2}).$$

The assumed Bernstein–von Mises limit and the delta method give the stated posterior limit for $\psi(\vartheta)$. The parametric information bound for regular estimators under the same correctly specified joint model gives

$$V_{2S} - D\psi(\vartheta_0)\mathcal{I}(\vartheta_0)^{-1}D\psi(\vartheta_0)' \succeq 0.$$

For a regular asymptotically linear estimator, equality holds if and only if its influence function equals the efficient influence function almost surely. $\square$

APPENDIX S3. FULL SETS OF IMPULSE RESPONSES

Figures [S1-S4](#) display the full set of impulse responses for the selected 13-variable system for the boom and bust application. Figures [S5-S8](#) display the full set of replicated impulse responses for the original 10-variable heteroskedastic system of BPSS. The replication uses updated monthly data and the monthly Sims-Zha prior, which is similar to the prior used by BPSS. To be compatible with BPSS’s original article, the scale of impulse responses is not adjusted to represent percent or percentage points.

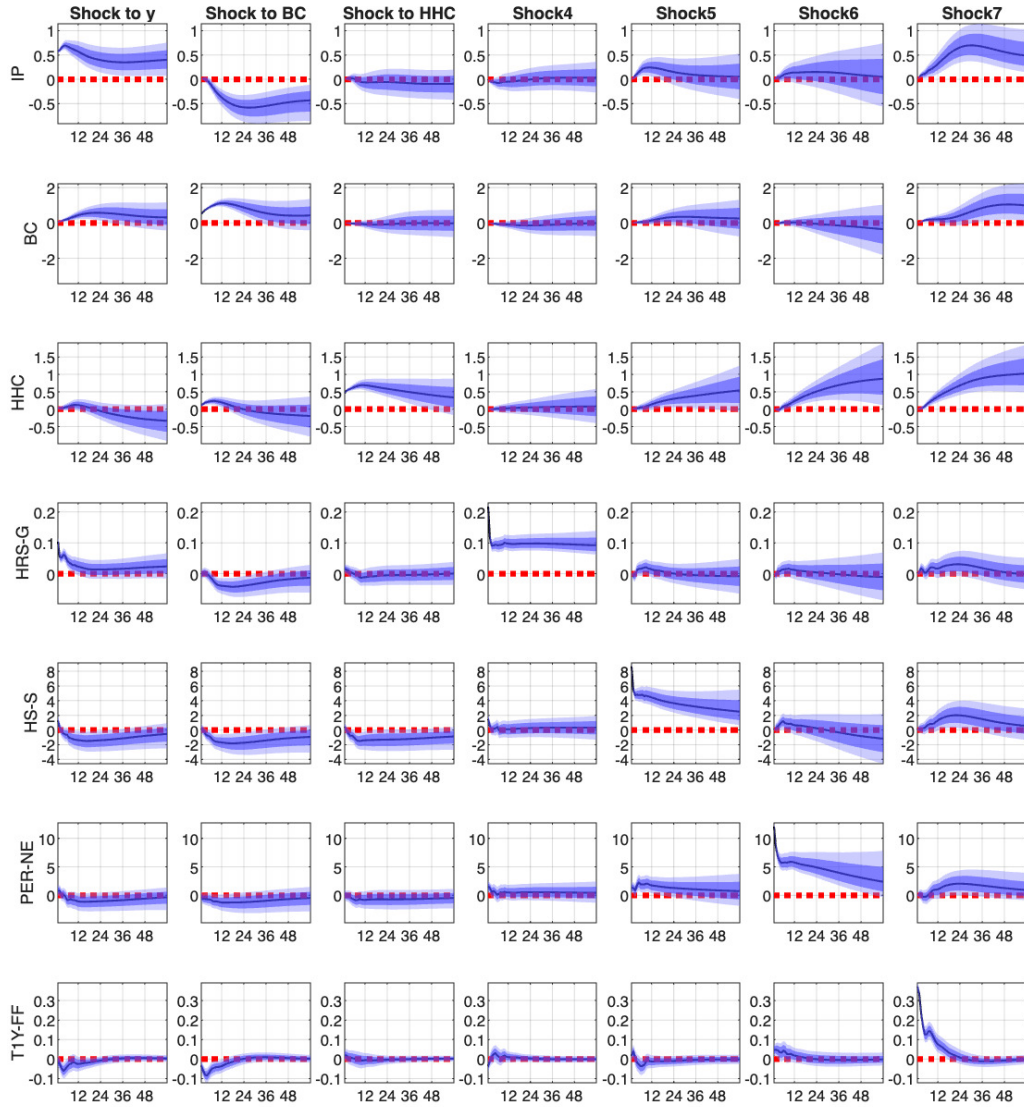


FIGURE S1. Part 1 of the full set of impulse responses for the selected 13-variable system. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. Table 1 provides complete descriptions of the variable labels shown on the y-axis.

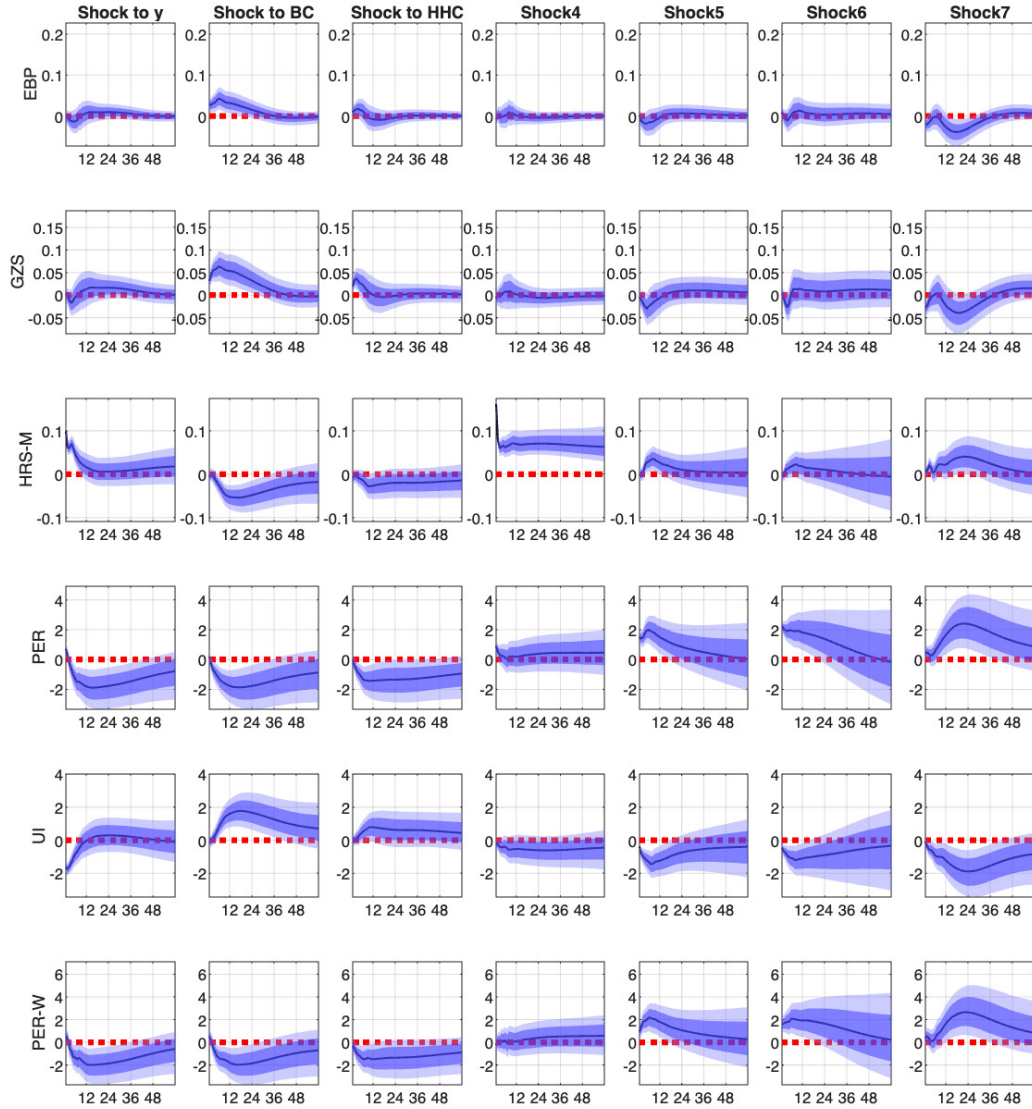


FIGURE S2. Part 2 of the full set of impulse responses for the selected 13-variable system. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. Table 1 provides complete descriptions of the variable labels shown on the y-axis.

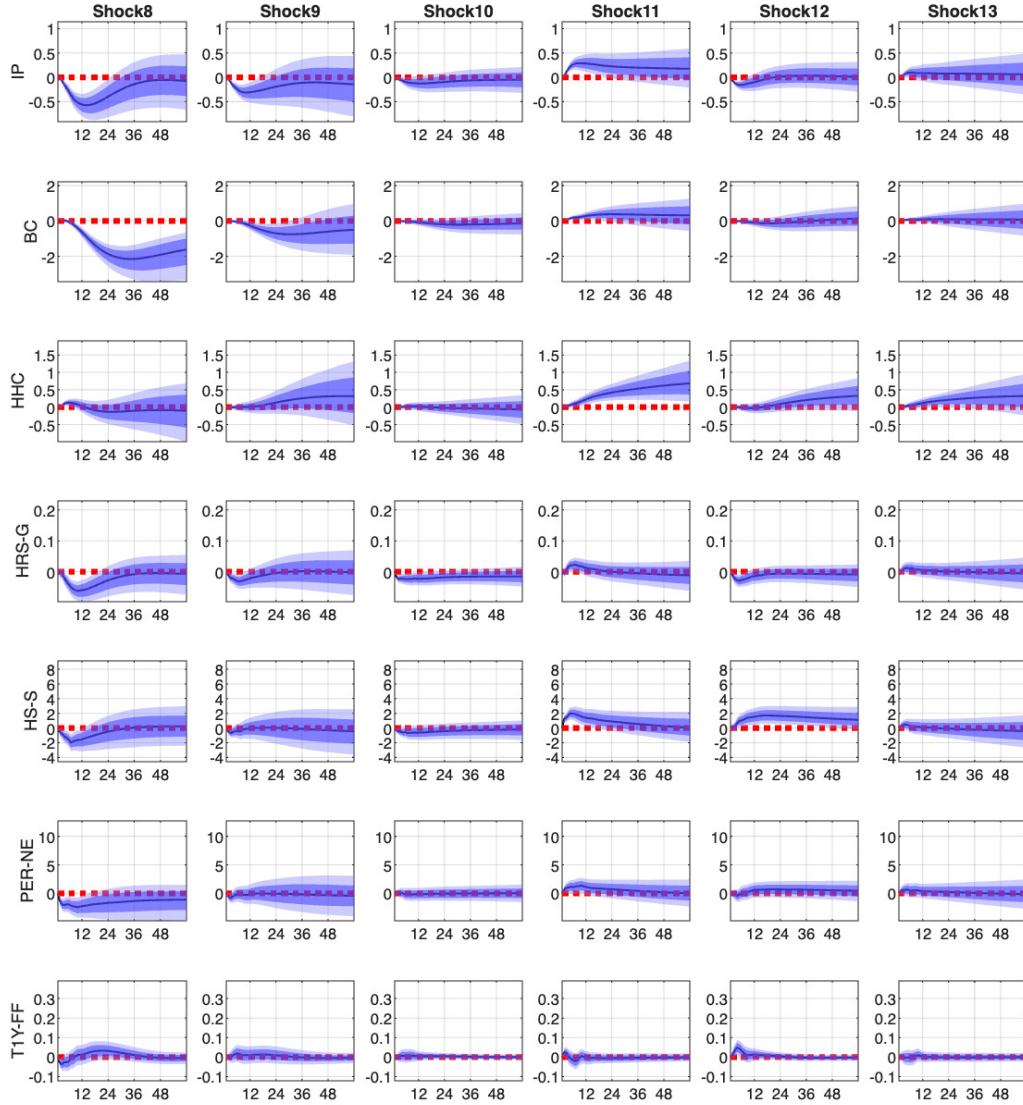


FIGURE S3. Part 3 of the full set of impulse responses for the selected 13-variable system. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. Table 1 provides complete descriptions of the variable labels shown on the y-axis.

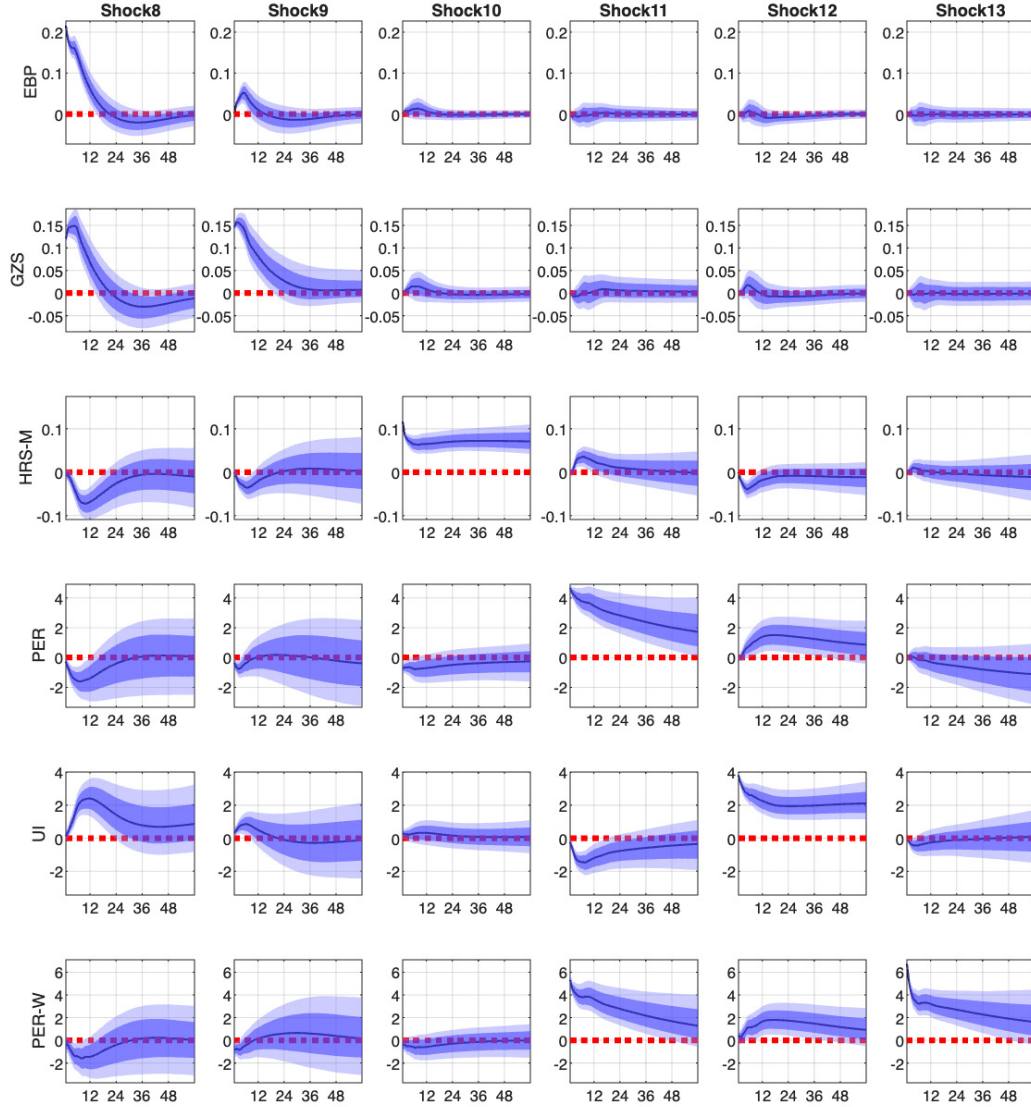


FIGURE S4. Part 4 of the full set of impulse responses for the selected 13-variable system. All impulse responses are expressed in percent. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. Table 1 provides complete descriptions of the variable labels shown on the y-axis.

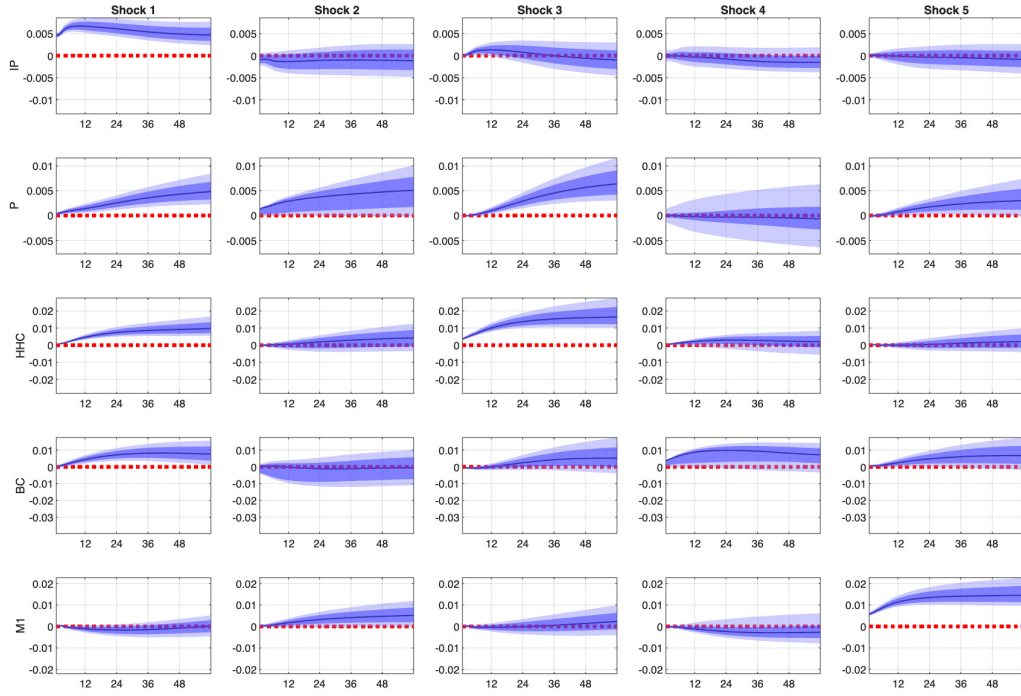


FIGURE S5. Part 1 of the full set of impulse responses for the original 10-variable heteroskedastic system of BPSS. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. The y-axis labels follow BPSS, who provide complete variable descriptions.

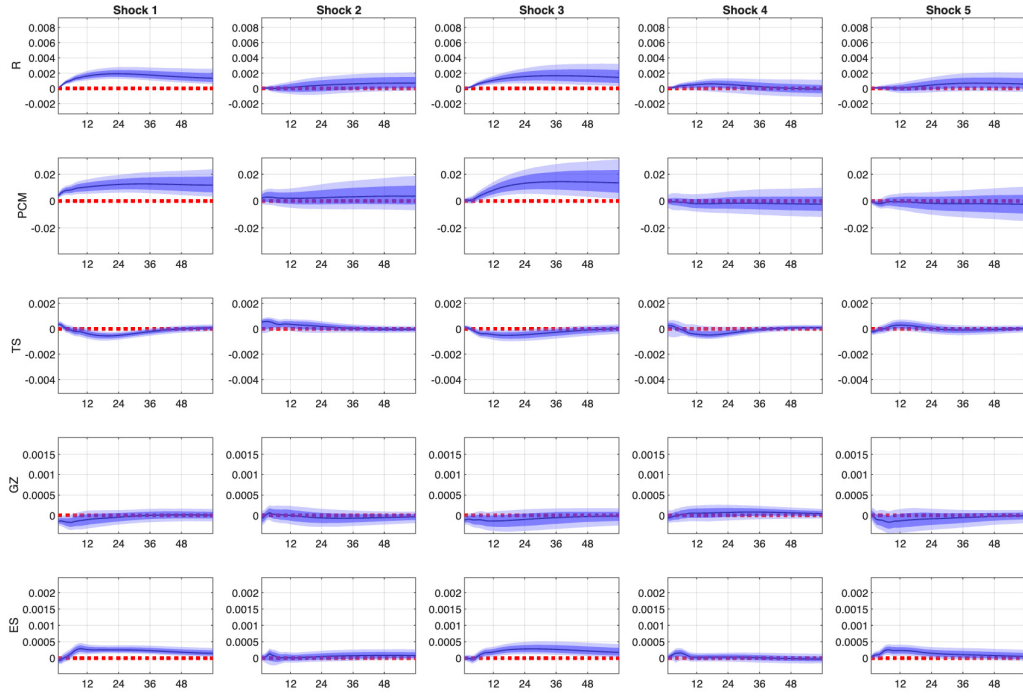


FIGURE S6. Part 2 of the full set of impulse responses for the original 10-variable heteroskedastic system of BPSS. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. The y-axis labels follow BPSS, who provide complete variable descriptions.

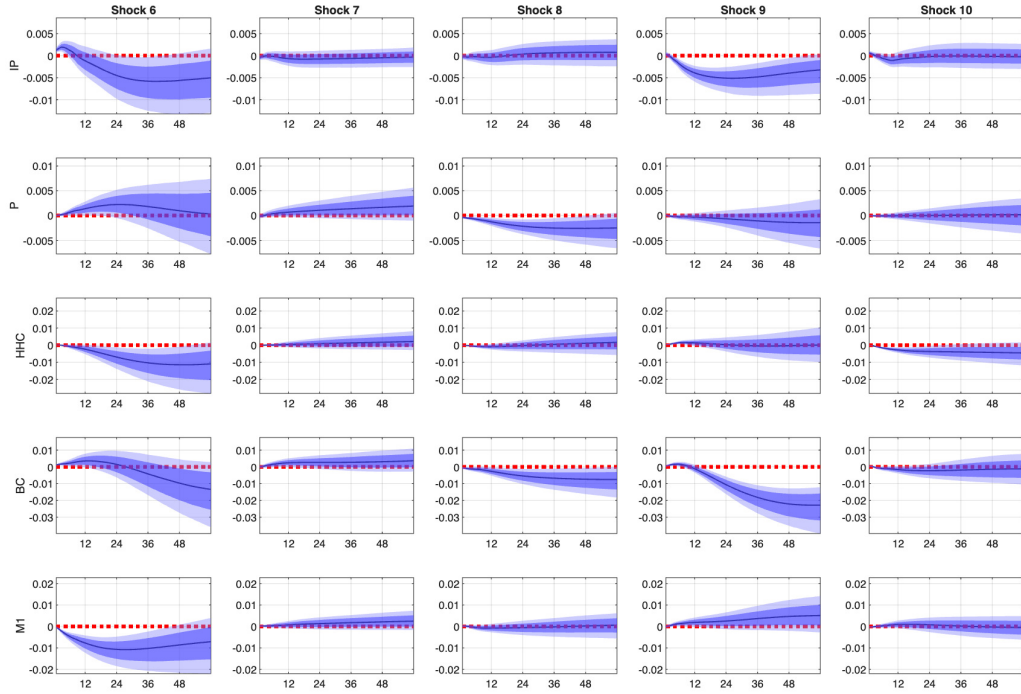


FIGURE S7. Part 3 of the full set of impulse responses for the original 10-variable heteroskedastic system of BPSS. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. The y-axis labels follow BPSS, who provide complete variable descriptions.

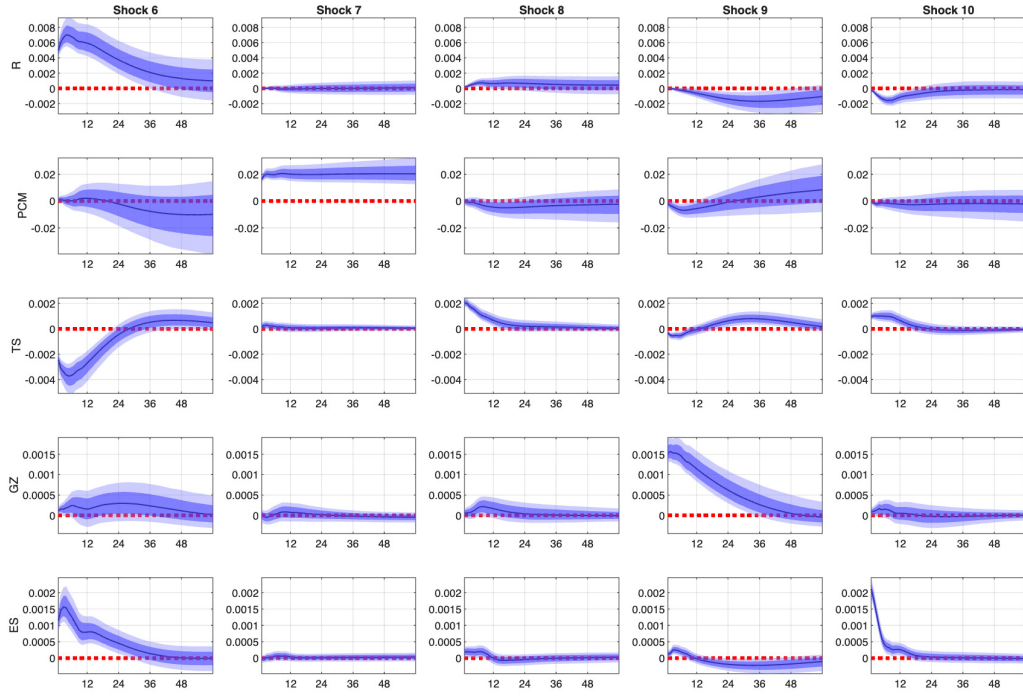


FIGURE S8. Part 4 of the full set of impulse responses for the original 10-variable heteroskedastic system of BPSS. The dark band shows the 68% posterior credible band and the light band the 90% posterior credible band. The y-axis labels follow BPSS, who provide complete variable descriptions.

REFERENCES

- Jonas E. Arias, Juan F. Rubio-Ramírez, and Daniel F. Waggoner. Inference in Bayesian proxy-SVARs. *Journal of Econometrics*, 225(1):88–106, 2021. doi: 10.1016/j.jeconom.2020.12.004.
- Ben S. Bernanke, Jean Boivin, and Piotr Elias. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1):387–422, 2005.
- Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.
- Markus K. Brunnermeier, Darius Palia, Karthik A. Sastry, and Christopher A. Sims. Feedbacks: Financial markets and economic activity. *American Economic Review*, 111:1845–1879, 2021. DOI: 10.1257/aer.20180733.
- Dario Caldara and Edward Herbst. Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs. *American Economic Journal: Macroeconomics*, 11(1):157–192, 2019. doi: 10.1257/mac.20170294.
- Lawrence J. Christiano, Martin S. Eichenbaum, and Charles L. Evans. Monetary policy shocks: What have we learned and to what end? In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1A, pages 65–148. North-Holland, Amsterdam, Holland, 1999.
- Morris A. Davis and Jonathan Heathcote. Housing and the business cycle. *International Economic Review*, 46(3):751–784, 2005. doi: 10.1111/j.1468-2354.2005.00345.x.
- Mario Forni and Luca Gambetti. Sufficient information in structural VARs. *Journal of Monetary Economics*, 66:124–136, 2014. doi: 10.1016/j.jmoneco.2014.04.005.
- Mark Gertler and Peter Karadi. Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76, 2015. doi: 10.1257/mac.20130329. URL <http://dx.doi.org/10.1257/mac.20130329>.
- Domenico Giannone and Lucrezia Reichlin. Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2–3):455–465, 2006. doi: 10.1162/jeea.2006.4.2-3.455.
- Domenico Giannone, Michele Lenza, and Giorgio E. Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2401–2436, 2021.
- Simon Gilchrist and Egon Zakrajšek. Credit supply shocks and economic activity in a financial accelerator model. In Alan Blinder, Andrew Lo, and Robert Solow, editors, *Rethinking the Financial Crisis*. Russell Sage Foundation, 2012.
- Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.
- João Gomes, Urban Jermann, and Lukas Schmid. Sticky leverage. *American Economic Review*, 106(12):3800–3828, 2016. doi: 10.1257/aer.20130952.

- Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273, 2020.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2 edition, 2009.
- Marek Jarociński and Bartosz Maćkowiak. Granger causal priority and choice of variables in vector autoregressions. *Review of Economics and Statistics*, 99(2):319–329, 2017. doi: 10.1162/REST_a_00601.
- Edward E. Leamer. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. John Wiley & Sons, New York, 1978.
- Edward E. Leamer. Housing is the business cycle. In *Housing, Housing Finance, and Monetary Policy*, Proceedings, Economic Policy Symposium, Jackson Hole, pages 149–233. Federal Reserve Bank of Kansas City, 2007.
- Edward E. Leamer. Housing really is the business cycle: What survives the lessons of 2008–09? *Journal of Money, Credit and Banking*, 47(S1):43–50, 2015.
- Karel Mertens and Morten O. Ravn. The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review*, 103(4):1212–1247, 2013. doi: 10.1257/aer.103.4.1212.
- Atif Mian, Amir Sufi, and Emil Verner. Household debt and business cycles worldwide. *The Quarterly Journal of Economics*, 132(4):1755–1817, 2017. doi: 10.1093/qje/qjx017.
- Karsten Müller and Emil Verner. Credit allocation and macroeconomic fluctuations. *The Review of Economic Studies*, 91(6):3645–3676, November 2024. doi: 10.1093/restud/rdad112.
- Berardino Palazzo and Ram Yamarthy. Credit risk and the transmission of interest rate shocks. *Journal of Monetary Economics*, 130:120–136, 2022. doi: 10.1016/j.jmoneco.2022.06.004.
- Lucrezia Reichlin. Dynamic factor models for big data, 26 years later: what have we learned? Walras-Bowley Lecture, North American Summer Meeting of Econometrics Society, Atlanta, June 2026.
- Matthew Rognlie, Andrei Shleifer, and Alp Simsek. Investment hangover and the great recession. *American Economic Journal: Macroeconomics*, 10(2):113–153, April 2018. doi: 10.1257/mac.20160211.
- Juan Francisco Rubio-Ramírez, Daniel F. Waggoner, and Tao Zha. Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77:665–696, 2010.
- Christopher A. Sims and Tao Zha. Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968, 1998.
- Christopher A. Sims and Tao Zha. Error bands for impulse responses. *Econometrica*, 67(5):1113–1155, 1999.
- Christopher A. Sims and Tao Zha. Does monetary policy generate recessions? *Macroeconomic Dynamics*, 10:231–272, 2006.

- James H. Stock and Mark W. Watson. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610): 917–948, 2018. doi: 10.1111/ecoj.12593.
- Eric T. Swanson. Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, 118:32–53, 2021. doi: 10.1016/j.jmoneco.2020.09.013.
- A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. doi: 10.1017/CBO9780511802256.
- Daniel F. Waggoner and Tao Zha. A gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, 28(2):349–366, 2003.
- Tao Zha. Block recursion and structural vector autoregressions. *Journal of Econometrics*, 90(2):291–316, June 1999.