BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2025-25

Last updated: September 19, 2025

Financial Shocks and the Output Growth Distribution

François-Michel Boire
Department of Mathematics and Statistics
University of Ottawa
francois-michel.boire@uottawa.ca

Thibaut Duprey
Financial Stability Department
Bank of Canada
tduprey@bankofcanada.ca

Alexander Ueberfeldt
Canadian Economic Analysis Department
Bank of Canada
aueberfeldt@bankofcanada.ca

DOI: https://doi.org/10.34989/swp-2025-25 | ISSN 1701-9397

Acknowledgements

We thank Paul Beaudry, Gabriel Bruneau, Heng Chen, Bruno Feunou, Sermin Gungor, Ruben Hipp, Gary Koop, Stéphane Surprenant, Kerem Tuzcuoglu and seminar participants at the Bank of Canada.

Abstract

This paper studies how financial shocks shape the distribution of output growth by introducing a quantile-augmented vector autoregression (QAVAR), which integrates quantile regressions into a structural VAR framework. The QAVAR preserves standard shock identification while delivering flexible, nonparametric forecasts of conditional moments and tail risk measures for gross domestic product (GDP). Applying the model to financial conditions and credit spread shocks, we find that adverse financial shocks worsen the downside risk to GDP growth significantly, while the median and upper percentiles respond more moderately. This underscores the importance of nonlinearities and heterogeneous tail dynamics in assessing macro-financial risks.

Topics: Central bank research, Econometric and statistical methods, Financial markets, Financial stability, Monetary and financial indicators

JEL codes: C32, C53, E32, E44, G01

Résumé

Dans cette étude, nous examinons comment les chocs financiers influent sur la distribution de la croissance de la production. Nous introduisons un modèle vectoriel autorégressif semi quantile qui intègre des régressions quantiles à un cadre vectoriel autorégressif structurel. Ce modèle conserve l'identification conventionnelle des chocs tout en permettant de produire des prévisions flexibles et non paramétriques des moments conditionnels ainsi que des mesures de risque extrême pour le produit intérieur brut (PIB). En appliquant le modèle aux conditions financières et aux chocs d'écarts de crédit, nous constatons que les chocs financiers défavorables accentuent considérablement les risques à la baisse entourant la croissance du PIB, tandis que les centiles médian et supérieurs de la croissance du PIB présentent des réactions plus modérées. Ces constats soulignent l'importance des non-linéarités et de la dynamique hétérogène des queues de distribution dans l'évaluation des risques macrofinanciers.

Sujets : Recherches menées par les banques centrales, Méthodes économétriques et statistiques, Marchés financiers, Stabilité financière, Indicateurs monétaires et financiers
Codes JEL : C32, C53, E32, E44, G01

1 Introduction

Financial shocks are a major source of macroeconomic fluctuations, but their effects vary with economic conditions. At times, adverse shocks to financial markets coincide with mild slowdowns; in other cases, they precipitate deep recessions. Identifying when financial shocks translate into material downside risk to growth is essential for both policymakers and market participants, particularly in periods of elevated uncertainty or fragility.

This paper studies how financial shocks affect the distribution of gross domestic product (GDP) growth. We first develop a quantile-augmented vector autoregressive model (QAVAR) by incorporating regressed conditional quantiles into a VAR while preserving conventional shock identification. We then apply the QAVAR to revisit two seminal studies—Adrian et al. (2019) and Gilchrist and Zakrajšek (2012)—on the effects of financial shocks on GDP growth.

The QAVAR is an extension of the standard VAR that retains conventional shock identification while recovering the conditional distribution of a single variable—GDP growth in our case. By numerically integrating over a finite set of fitted quantiles, the model yields flexible density forecasts and direct estimates of conditional moments and risk measures, without imposing a parametric distribution.

Our empirical application unifies the strengths of Gilchrist and Zakrajšek (2012) (hereafter GZ), who identify structural financial shocks but do not examine distributional effects, and Adrian et al. (2019) (hereafter ABG), who study distributional impacts of financial conditions without structural identification. We embed ABG's setup in our QAVAR and show that structural financial conditions shocks worsen downside risks to GDP growth, as measured by the 5th percentile (Growth-at-Risk). We then extend GZ's setup and find that corporate bond market stress worsens downside risk: it is twice as detrimental as the average effect in GZ's model. In contrast, the median and upper percentiles of the GDP response to excess bond premium shocks are about 30% less severe than the average effect in GZ's linear model.

Our findings align with the literature showing that financial turmoil deepens recessions and slows recoveries (e.g., Reinhart and Rogoff, 2014). They also support Loria et al. (2024), who find that the 10th percentile of GDP growth falls more sharply than the median or 90th percentile after financial shocks. Unlike their quantile local projection approach, our method allows for both shock identification and dynamic quantile evolution.

We show that our QAVAR identifies similar structural shocks to those in the linear VAR of GZ. Monte Carlo simulations confirm that quantile-based conditional moment estimates perform well in settings with volatility clustering and non-Gaussian innovations. The idea of

using a sequence of quantiles to approximate distributions and simulate non-linear dynamics is not new. In recent years, many applications have drawn from the same heuristic projection to uncover general model-free distribution effects.[1]

Our QAVAR offers a middle ground between standard VARs and fully quantile-based systems by modelling the distribution of a single variable. Models such as the conditional autoregressive VAR (CAViaR) (Engle and Manganelli, 2004; White et al., 2015), cast in a VAR framework by Chavleishvili et al. (2021) and Chavleishvili and Manganelli (2024), model the entire system in quantiles and use quantile shocks to simulate tail scenarios. Instead, our approach can retain standard VAR structural shock identification while capturing relevant nonlinear propagation and tail dynamics.

The rest of the paper is organized as follows. Section 2 outlines the QAVAR modelling framework and outlines impulse response function simulation procedures. Section 3 revisits the works of Adrian et al. (2019) and Gilchrist and Zakrajšek (2012) to highlight the contribution of non-linear dynamics in the propagation of structural financial shocks. Section 4 concludes. Monte Carlo simulations comparing the conditional moments estimated by the QAVAR to those of the standard VAR counterpart are left for the appendix.

2 The model

This section first presents the standard VAR model before introducing the QAVAR model. The key difference in the QAVAR is that a single variable of interest relies on a quantile-based density forecast while the other variables are modelled as in a VAR. To ease exposition, we discuss the case of a bivariate (QA)VAR with one lag, but the framework easily extends to general multivariate autoregressive vectors with extended lag structures. This flexible approach allows for non-linear transmission of shocks through heterogeneous conditional quantile responses of the variable of interest, while preserving the structural identification methods of a standard VAR. Because quantile regression (QR) estimates are distribution-free, the hybrid model does not yield analytically tractable forecasts. Hence, we rely on density forecasts to simulate the mean and quantile impulse response functions.

2.1 The standard VAR

Let $\{Y_t = (Y_{1,t}, Y_{2,t})' : t = 0, 1, \ldots, T\}$ denote a weakly stationary bivariate process governed by a first-order VAR, $Y_t = \mu + \mathbf{B}Y_{t-1} + \epsilon_t$, where $\mu = (\mu_1, \mu_2)$ is a vector of intercepts, $\mathbf{B}$ is

[1] Quantile-based density estimates have been implemented, for instance, within deep Q-learning algorithms (see, e.g., Bellemare et al., 2017; Dabney et al., 2018). Other relevant applications include stock returns forecasting (Arısoy, 2023) and multi-objective hedging (Cao et al., 2020), to name only a few.

a matrix of autoregressive coefficients satisfying the usual stationarity conditions, and $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ is a vector of serially independent innovations satisfying $\mathbb{E}[\epsilon_t] = \mathbf{0}$ and $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathbf{\Sigma}$. In matrix form, we can write

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} . \tag{1}$$

In the first-order VAR, the expectation of Y_t given Y_{t-1}, denoted $\overline{Y}_t = \mathbb{E}_{t-1}[Y_t]$, is given by $\overline{Y}_t = (\overline{Y}_{1,t}, \overline{Y}_{2,t})' = \mu + \mathbf{B} Y_{t-1}$, and the system in Equation (1) can be formulated as

$$Y_t = \overline{Y}_t + \epsilon_t \, , \tag{2}$$

where $\overline{Y}_t$ represents the mean forecast of Y_t given Y_{t-1}, and ϵ_t captures the innovation component.

The formulation in Equation (2) highlights the decomposition of each realization of Y_t into a model-implied predictable component and a corresponding mean-zero innovation. In the VAR above, the predictable component is a linear function of lagged variables where intercept and autoregressive coefficient estimates are obtained by ordinary least-squares (OLS) and denoted $\hat{\mu}^{\text{OLS}}$ and $\hat{\mathbf{B}}^{\text{OLS}}$, respectively. Under the usual OLS assumptions, the fitted value denoted with a hat, $\widehat{\overline{Y}}_t = \hat{\mu}^{\text{OLS}} + \hat{\mathbf{B}}^{\text{OLS}} Y_{t-1}$, is the best linear unbiased estimator of $\overline{Y}_t$, and innovations are then identified as regression residuals $\hat{\epsilon}_t = (\hat{\epsilon}_{1,t}, \hat{\epsilon}_{2,t}) = Y_t - \widehat{\overline{Y}}_t$. Importantly, this decomposition into predictable (fitted) and unpredictable (residual) values accommodates alternative models and hence other estimators of $\mathbb{E}_{t-1}[Y_t]$. Our model novelty is to replace $\overline{Y}_t$ with a mean forecast based on conditional quantile estimates.

2.2 The quantile-augmented VAR

The QAVAR modifies the standard linear VAR framework in Equation (2) to capture state-dependent distribution features beyond the first moment. One variable, $Y_{1,t}$, is modelled using conditional quantile estimates. The resulting model-free density estimates, unlike the linear VAR, capture time-varying higher-order moments and nonlinear shock transmission. The other variable, $Y_{2,t}$, is still modelled linearly as in a standard VAR, such that $Y_{2,t}|Y_{t-1}$ maintains the same linear form. Any additional variable in the QAVAR would continue to follow the linear VAR framework; thus, the case with more than two variables is a simple extension of the bivariate setup presented here.

Define $F_{Y_{1,t}|Y_{t-1}}$ as the conditional cumulative distribution function (CDF) of $Y_{1,t}$ given Y_{t-1} and its generalized inverse (or quantile function) as $Q_{Y_{1,t}|Y_{t-1}}$. Moreover, quantile residu-

als for each quantile $\tau \in (0,1)$ are defined as $\eta_{t,\tau} = Y_{1,t} - Q_{Y_{1,t}|Y_{t-1}}(\tau)$, satisfying $Q_{\eta_{t,\tau}}(\tau) = 0$ by construction (Koenker and Bassett Jr, 1978).

Next, consider a partition of the probability space given by an evenly spaced grid, $\mathcal{T}_N = \{\tau_n = n/(N+1) : n = 1, \ldots, N\}$.[2] The quantile-based approximation of $\mathbb{E}_{t-1}[Y_{1,t}]$ is then given by

$$\widetilde{Y}_{1,t}^{(N)} = \frac{1}{N} \sum_{\tau \in \mathcal{T}_N} Q_{Y_{1,t}|Y_{t-1}}(\tau) \, . \tag{3}$$

Under mild integrability assumptions, the approximation $\widetilde{Y}_{1,t}^{(N)}$ converges to $\int_0^1 Q_{Y_{1,t}|Y_{t-1}}(\tau)\, d\tau = \mathbb{E}_{t-1}[Y_{1,t}]$ as $N \to \infty$. Beyond the conditional mean, this simple numerical integration approach lends itself to the approximation of other relevant risk metrics—such as conditional variance, skewness, kurtosis, value-at-risk, expected shortfall, and more—via appropriate transformations of the quantile function. More generally, this quantile-based fit is designed to capture heterogeneous effects across quantiles of the conditional distribution $Y_{1,t}|Y_{t-1}$.

We modify Equation (2) to incorporate $\widetilde{Y}_{1,t}^{(N)}$ into the VAR framework. Specifically, we replace the traditional conditional expectation $\overline{Y}_{1,t}$ with $\widetilde{Y}_{1,t}^{(N)}$, the quantile-based approximation of $\mathbb{E}_{t-1}[Y_{1,t}]$ representing the predictable component in the QAVAR model. Here, the tilde denotes a quantile-based conditional expectation, while the bar indicates the conventional linear conditional expectation. The resulting system is written as

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \widetilde{Y}_{1,t}^{(N)} \\ \overline{Y}_{2,t} \end{bmatrix} + \begin{bmatrix} \widetilde{\epsilon}_{1,t}^{(N)} \\ \epsilon_{2,t} \end{bmatrix}, \tag{4}$$

where the (approximated) forecast error for $Y_{1,t}$ is given by $\widetilde{\epsilon}_{1,t}^{(N)} = Y_{1,t} - \widetilde{Y}_{1,t}^{(N)} = \frac{1}{N} \sum_{\tau \in \mathcal{T}_N} \eta_{t,\tau}$. This expression captures the deviation between the realized value of $Y_{1,t}$ and its quantile-based approximation constructed from N conditional quantiles. By the same numerical integration argument used in Equation (3), the fitted QAVAR residual $\widetilde{\epsilon}_t^{(N)} = (\widetilde{\epsilon}_{1,t}^{(N)}, \epsilon_{2,t})'$ from Equation (4) converges to the model-implied forecast error $\widetilde{\epsilon}_t^{*} = (\widetilde{\epsilon}_{1,t}^{*}, \epsilon_{2,t})' = Y_t - \mathbb{E}_{t-1}[Y_t]$ as $N \to \infty$.

In summary, the system in Equation (4) approximates the true data-generating process specifying the quantile function $Q_{Y_{1,t}|Y_{t-1}}(\tau)$ across all $\tau \in (0,1)$. To incorporate QR into the VAR framework, we adopt a hybrid approach whereby only a finite collection of quantiles are considered for $Y_{1,t}$, while the second variable, $Y_{2,t}$, remains governed by its conventional

[2]The appropriate number of quantiles N should be proportional to the sample size T, since larger T permits estimation of quantiles further into the distribution's tails. In the Monte Carlo simulations reported in Appendix C, we set $N = 99$ and examine results for $T = 200$, 500, and 1000.

linear VAR form.[3]

2.3 Estimating conditional expectations with quantile regressions

The hybrid QAVAR model approximates the conditional expectation in Equation (3) by averaging N fitted quantiles of $Y_{1,t}|Y_{t-1}$. We now detail how to compute those quantile estimates. Assuming a first-order autoregressive structure as above, we see that the conditional quantile estimator follows a linear specification given by

$$Q_{Y_{1,t}|Y_{t-1}}(\tau) = \alpha_\tau + \boldsymbol{\beta}_\tau Y_{t-1} = \alpha_\tau + \beta_{1,1,\tau} Y_{1,t-1} + \beta_{1,2,\tau} Y_{2,t-1} \ , \tag{5}$$

where the coefficients α_τ and $\boldsymbol{\beta}_\tau = (\beta_{1,1,\tau}, \beta_{1,2,\tau})$ are indexed by the quantile level τ. QR estimates $\hat{\alpha}_\tau$ and $\hat{\boldsymbol{\beta}}_\tau = (\hat{\beta}_{1,1,\tau}, \hat{\beta}_{1,2,\tau})$ are obtained by solving

$$\hat{\alpha}_\tau, \hat{\boldsymbol{\beta}}_\tau = \underset{\alpha,\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_\tau \left(Y_{1,t} - \alpha - \boldsymbol{\beta} Y_{t-1} \right) \ , \tag{6}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}_{\{u<0\}})$ is the standard check function (Koenker and Bassett Jr, 1978). The estimated autoregressive coefficients across quantiles are stored in the matrices

$$\hat{\alpha}^{\text{QR}} = \begin{bmatrix} \hat{\alpha}_{\tau_1} \\ \vdots \\ \hat{\alpha}_{\tau_N} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{B}}^{\text{QR}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\tau_1} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{\tau_N} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{1,1,\tau_1} & \hat{\beta}_{1,2,\tau_1} \\ \vdots & \vdots \\ \hat{\beta}_{1,1,\tau_N} & \hat{\beta}_{1,2,\tau_N} \end{bmatrix} \ .$$

Note that for each $\tau \in \mathcal{T}_N$, the QR estimation is carried out independently. Hence, the rows of $\hat{\alpha}^{\text{QR}}$ and $\hat{\mathbf{B}}^{\text{QR}}$ are estimated separately.[4]

If QR coefficients are constant across τ, the model reduces to a standard VAR. Otherwise, the heterogeneity in coefficients across quantiles captures non-linear and asymmetric

[3] Extensions to multivariate QR lie beyond the scope of this analysis, as such generalizations would necessitate accounting for dependence across conditional quantile functions of different variables. We instead assume that the conditional distribution of $Y_{1,t}$ is a function of past innovations alone, just like the rest of the system. A possible solution to modelling multiple quantile processes is offered by Chavleishvili and Manganelli (2024), who assume a triangular factorization of structural quantile shocks. In other words, a given conditional quantile of one variable depends on the quantiles of preceding variables in a user-chosen causal ordering.

[4] We acknowledge the issue of quantile crossing with a linear specification, where fitted quantile functions may not be generally monotone in τ. This phenomenon is more prevalent in data-poor regions of the distribution's support domain and may arise more frequently for tail events, such as deep recessions. To compute tail risk metrics, we ensure monotonicity by simply sorting fitted quantiles (Chernozhukov et al., 2010). More computationally intensive solutions would include those proposed by He (1997) and Bondell et al. (2010), which enforce monotonicity through parameter restrictions, and by Schmidt and Zhu (2016), where the spacings between adjacent quantiles are modelled as positive functions.

responses of $Y_{1,t}$ to lagged values, allowing the model to approximate conditional higher moments and multimodality. QR thus provides a flexible, semi-parametric representation of the conditional distribution of $Y_{1,t}|Y_{t-1}$, relaxing the distributional assumptions typically required to estimate conditional moments.

Given the quantile approximation of the conditional mean $\widetilde{Y}_{1,t}^{(N)}$ in Equation (3), the estimated conditional mean is

$$\hat{\widetilde{Y}}_{1,t}^{(N)} = \frac{1}{N} \sum_{\tau \in \mathcal{T}} \hat{Q}_{Y_t|Y_{t-1}}(\tau) \, , \tag{7}$$

with corresponding residual $\hat{\widetilde{\epsilon}}_{1,t}^{(N)} = Y_t - \hat{\widetilde{Y}}_{1,t}^{(N)}$. To express the complete QAVAR system, define respectively the stacked estimated intercept terms and autoregressive coefficients as

$$\hat{\mu} = \begin{bmatrix} \hat{\alpha}^{\text{QR}} \\ \hdashline [\hat{\mu}^{\text{OLS}}]_{-1} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{\tau_1} \\ \vdots \\ \hat{\alpha}_{\tau_N} \\ \hdashline \hat{\mu}_2^{\text{OLS}} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{B}}^{\text{QR}} \\ \hdashline [\hat{\mathbf{B}}^{\text{OLS}}]_{-1} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{1,1,\tau_1} & \hat{\beta}_{1,2,\tau_1} \\ \vdots & \vdots \\ \hat{\beta}_{1,1,\tau_N} & \hat{\beta}_{1,2,\tau_N} \\ \hdashline \hat{\beta}_{2,1} & \hat{\beta}_{2,2} \end{bmatrix} \, ,$$

where $[\hat{\mu}^{\text{OLS}}]_{-1}$ denotes the vector $\hat{\mu}^{\text{OLS}}$ without the intercept for $Y_{1,t}$, and $[\hat{\mathbf{B}}^{\text{OLS}}]_{-1}$ denotes the matrix $\hat{\mathbf{B}}^{\text{OLS}}$ without the coefficients that correspond to $Y_{1,t}$, which is instead modelled via QR.

The vector of $N+1$ fitted values (N conditional quantile estimates for $Y_{1,t}$, plus one conditional mean estimate for $Y_{2,t}$) is then $\hat{\Lambda}_t = \hat{\mu} + \hat{\mathbf{B}} Y_{t-1} = (\hat{Q}_{Y_{1,t}|Y_{t-1}}(\tau_1), \ldots, \hat{Q}_{Y_{1,t}|Y_{t-1}}(\tau_N), \hat{\overline{Y}}_{2,t})'$, and the corresponding residual vector is $\hat{H}_t = (\hat{\eta}_{t,\tau_1}, \ldots, \hat{\eta}_{t,\tau_N}, \hat{\epsilon}_{2,t})'$. The QAVAR system is written as

$$Y_t = \mathbf{W}(\hat{\mu} + \hat{\mathbf{B}} Y_{t-1} + \hat{H}_t) = \mathbf{W}(\hat{\Lambda}_t + \hat{H}_t) \, , \tag{8}$$

where the weighing matrix

$$\mathbf{W} = \begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

aggregates the N quantile forecasts and the N quantile residuals for $Y_{1,t}$ and leaves the dynamics of other endogenous variables intact. This formulation highlights the QAVAR model's ability to incorporate heterogeneous conditional dynamics and non-linear propagation of shocks in a simple VAR-style representation given by

$$Y_t = \hat{\widetilde{\mu}} + \hat{\widetilde{\mathbf{B}}} Y_{t-1} + \hat{\widetilde{\epsilon}}_t^{(N)} = \hat{\widetilde{\Lambda}}_t + \hat{\widetilde{\epsilon}}_t^{(N)} \, , \tag{9}$$

where $\hat{\tilde{\mu}} = \mathbf{W}\hat{\mu}$, $\hat{\tilde{\mathbf{B}}} = \mathbf{W}\hat{\mathbf{B}}$, and $\hat{\tilde{\epsilon}}_t^{(N)} = (\hat{\tilde{\epsilon}}_{1,t}^{(N)}, \hat{\epsilon}_{2,t})' = \mathbf{W}\hat{H}_t$. Note that the QAVAR forecast $\hat{\tilde{\Lambda}}_t$ is analogous to the one derived in Equation (2) for the VAR. Altogether, the aggregated model in Equation (9) resembles a conventional VAR, while preserving a model-free quantile structure.

In summary, the key distinction between the VAR and QAVAR lies in the use of quantile aggregation to estimate the conditional expectation of $Y_{1,t}$. Beyond that, the two models share a common structure, and standard structural identification methods apply directly to the QAVAR residual vector $\hat{\tilde{\epsilon}}_t^{(N)}$.

2.4 Structural identification

Let the reduced-form shocks $\tilde{\epsilon}_t^* = (\tilde{\epsilon}_{1,t}^*, \epsilon_{2,1})'$ be expressed as a linear combination of orthogonal structural shocks $U_t = (U_{1,t}, U_{2,t})'$ with variance-covariance matrix $\mathbb{E}[U_t U_t'] = \mathbf{\Omega}$. Specifically, we assume $\tilde{\epsilon}_t^* = \mathbf{A}U_t$, where $\mathbf{A}$ is a full-rank square matrix. Consequently, the variance-covariance matrix of reduced-form shocks is given by $\mathbb{E}[\tilde{\epsilon}_t^* \tilde{\epsilon}_t^{*\prime}] = \tilde{\mathbf{\Sigma}} = \mathbf{A}\mathbf{\Omega}\mathbf{A}'$. Imposing identifying restrictions on the entries of the matrix $\mathbf{A}$ (via, e.g., Cholesky-based orderings and sign/zero restrictions) yields a structural representation that permits economically meaningful interpretation of the vector U_t, consistent with the conventional structural VAR.

In the context of the QAVAR, reduced-form shocks are approximated as

$$\tilde{\epsilon}_t^{(N)} = \mathbf{A}^{(N)} U_t^{(N)} , \tag{10}$$

where $\mathbb{E}[U_t^{(N)} U_t^{(N)\prime}] = \mathbf{\Omega}^{(N)}$. Accordingly, the reduced-form variance-covariance matrix is approximated as $\mathbb{E}[\tilde{\epsilon}_t^{(N)} \tilde{\epsilon}_t^{(N)\prime}] = \tilde{\mathbf{\Sigma}}^{(N)} = \mathbf{A}^{(N)} \mathbf{\Omega}^{(N)} \mathbf{A}^{(N)\prime}$. Equation (10) highlights that structural identification in the QAVAR depends on the choice of the quantile grid $\mathcal{T}_N$ used to construct the reduced-form shock approximation $\tilde{\epsilon}_{1,t}^{(N)}$. If the grid is too coarse (i.e., small N), then the approximation error may interfere with the identification procedure. However, under regularity conditions on the QR estimators, we have that as $N \to \infty$, $\tilde{\epsilon}_t^{(N)} \to \tilde{\epsilon}_t^*$ and $U_t^{(N)} \to U_t$, implying convergence of the moments $\tilde{\mathbf{\Sigma}}^{(N)} \to \tilde{\mathbf{\Sigma}}$, $\mathbf{\Omega}^{(N)} \to \mathbf{\Omega}$, and $\mathbf{A}^{(N)} \to \mathbf{A}$. Consequently, the structural QAVAR becomes asymptotically equivalent to the standard structural VAR in terms of identification.[5] The estimated counterparts of the theoretical approximations $U_t^{(N)}, \mathbf{\Omega}^{(N)}, \tilde{\mathbf{\Sigma}}^{(N)}$ and $\mathbf{A}^{(N)}$ are denoted by $\hat{U}_t^{(N)}, \hat{\mathbf{\Omega}}^{(N)}, \hat{\tilde{\mathbf{\Sigma}}}^{(N)}$, and $\hat{\mathbf{A}}^{(N)}$, re-

[5] In practice, we find that using a moderately fine grid, such as $N = 99$, is sufficient for the quantile-based residuals and shock decompositions to closely replicate those obtained in the traditional VAR. Note that standard QR procedures are not computationally restrictive for moderate sample and model sizes, which allows the user to increase N as needed, at little additional cost.

spectively.

To integrate QR outputs into the structural VAR framework, we decompose the quantile residuals $\eta_{t,\tau}$ into a quantile-invariant structural component $\widetilde{\epsilon}_{1,t}^{(N)}$ and a quantile-specific disturbance $\zeta_{t,\tau}^{(N)}$ indexed by τ. For notational simplicity, we hereafter omit the superscript (N) and write

$$\eta_{t,\tau} = \widetilde{\epsilon}_{1,t} + \zeta_{t,\tau} \tag{11}$$

for all $\tau \in \mathcal{T}_N$. The analogous estimated relationship is given by $\hat{\eta}_{t,\tau} = \hat{\widetilde{\epsilon}}_{1,t} + \hat{\zeta}_{t,\tau}$. In this formulation, $\widetilde{\epsilon}_{1,t}$ captures a location shift of the mean forecast characterized by a homogeneous shock across quantiles, whereas $\zeta_{t,\tau}$ reflects heterogeneous effects changing the shape of the conditional distribution. Furthermore, note that the innovations $\hat{\widetilde{\epsilon}}_{1,t}$ are assumed i.i.d., implying that variations in the shape of the distribution are all attributed to $\hat{\zeta}_{t,\tau}$.

Substituting the decomposition of quantile residuals of Equation (11) in the residual vector $\hat{H}_t$ of Equation (8) and then applying the structural representation of Equation (10), we can rewrite Equation (8) as

$$Y_t = \mathbf{W}(\hat{\mu} + \hat{\mathbf{B}}Y_{t-1} + \hat{Z}_t) + \hat{\mathbf{A}}\hat{U}_t \ ,$$

where $\hat{Z}_t = (\hat{\zeta}_{t,\tau_1}, \ldots, \hat{\zeta}_{t,\tau_N}, 0)$ is a vector of quantile-specific disturbances. Since $\hat{Z}_t$ consists of deviations orthogonal to the structural shock, we have $\mathbf{W}\hat{Z}_t = \mathbf{0}$. Aggregating across quantiles then yields the quantile-augmented structural representation $Y_t = \hat{\widetilde{\Lambda}}_t + \hat{\mathbf{A}}\hat{U}_t$.

The key identifying assumption that allows the structural QAVAR representation is that, on impact, shocks change only the location of the conditional distribution without altering its shape. Subsequently, at a future time horizon, $k > 0$, the system's response to the shock propagates non-linearly through the term $\hat{\mathbf{B}}Y_{t+k-1} + \hat{Z}_{t+k}$, altering the shape of the distribution.

To illustrate, suppose that the one-step-ahead forecast of GDP growth at time t is normally distributed. At time t, an unexpected adverse financial shock occurs, leading to an immediate downward shift in all quantiles and preserving the Gaussian shape of the density. From time $t+1$ onward, however, the shock's effect is nonlinearly transmitted through heterogeneous quantile dynamics. As a result, the distribution of GDP growth may become skewed, fat-tailed, or even multimodal. Notably, the magnitude and shape of the response vary depending on the economy's initial condition. In particular, the adverse effects may be amplified under weak macroeconomic conditions, leading to a more pessimistic revision of future growth outlooks than what a standard VAR would predict.

In sum, while structural shocks in the QAVAR model produce a location shift on impact, subsequent propagation allows for non-linear responses across the conditional distribution of

one variable, extending the structural VAR toolkit to model hetereogenous quantile effects within a compact and estimable model.

2.5 Impulse response functions and simulation approach

A key feature of the structural VAR framework is its capacity to support counterfactual analysis, most notably through impulse response functions (IRFs). IRFs measure the dynamic response of an endogenous variable to an exogenous structural shock by comparing conditional trajectories under a baseline scenario and a counterfactual (or shocked) scenario. Although IRFs are conventionally used to characterize the evolution of the conditional mean, the same logic can be extended to study other distributional moments: for instance, to derive quantile impulse response functions (QIRF).

In contrast to linear VARs, non-linear VARs like the QAVAR allow for distributional IRFs that may depend on initial conditions or exhibit asymmetric or non-linear response to shocks (Gallant et al., 1993; Koop et al., 1996; Potter, 2000).[6] The nonlinear and distribution-free nature of the QAVAR framework complicates inference for distributional statistics.[7]

To address these challenges, we adopt a simulation-based approach to compute IRFs for various moments of the conditional distribution.[8] Specifically, we construct two measures: the mean IRF and the quantile IRF, which traces the one-step-ahead conditional quantile forecast at a given level $\tau \in \mathcal{T}_N$. These allow us to capture both conditional mean dynamics and tail risks.

Simulating the variable $Y_{1,t}$ modelled with QR requires an approximation of the conditional density of $Y_{1,t}|Y_{t-1}$, while so far we specified only its conditional quantiles. One method estimates the conditional density by interpolating and rescaling fitted quantiles (Schmidt and Zhu, 2016; Koenker, 2005). Alternatively, one can fit a parametric distribution to a set of quantiles (Adrian et al., 2019) or use kernel density estimates (Gaglianone and Lima, 2012; Davino et al., 2013; Korobilis, 2017). These methods are robust but can be computationally intensive. We instead adopt an efficient *ad-hoc* method by sampling directly from the set of N fitted conditional quantiles, sidestepping any issues related to estimating and sampling from an explicit density.

The simulation algorithm is outlined below.[9] The procedure concurrently simulates base-

[6]Note that in our QAVAR applications, we do not find a strong dependence to the magnitude or sign of the shock. This is not too surprising, since the shock propagates through a set of conditional linear quantiles, and the response to shocks is symmetric on impact (else for instance see Beaudry and Koop, 1993).

[7]The mean IRF in the QAVAR can be analytically tractable under suitable assumptions. See Appendix A for more details.

[8]Density forecast simulations by linearly weighting alternative forecasts have also become more popular in forecasting to improve accuracy, see Kapetanios et al. (2015).

[9]The same algorithm can be adapted to construct bootstrapped confidence intervals. See Appendix B for

line paths $\{Y_{t+k} : k = 0, 1, \ldots, h\}$ and shocked paths $\{Y^{\delta}_{t+k} : k = 0, 1, \ldots, h\}$ to construct IRF and QIRF estimators up to a user-chosen time horizon h.

1. **Initialization:** Given initial conditions Y_t and a structural shock δ_t, set $Y^{\delta}_t = Y_t + \hat{\mathbf{A}}\delta_t$. Sample a sequence of independent quantile levels $\{\theta_{t+1}, \ldots, \theta_{t+h}\}$ from $\mathcal{T}_N$ uniformly with probability $1/N$. Simulate a sequence of structural shocks $\{\hat{U}_{2,t+1}, \ldots, \hat{U}_{2,t+h}\}$ from an appropriate white noise process, such that $\hat{U}_{2,t} \sim \text{WN}(\mathbf{0}, [\hat{\tilde{\mathbf{\Omega}}}]_{-1})$, where $[\hat{\tilde{\mathbf{\Omega}}}]_{-1}$ is the sub-matrix of $\hat{\tilde{\mathbf{\Omega}}}$ that excludes covariance terms involving $\hat{U}_{1,t}$.

2. **Simulating the quantile-augmented variable:** For $0 < k \leq h$, we sample from a discrete approximation of the conditional densities of $Y_{1,t+k}|Y_{t+k-1}$ and $Y_{1,t+k}|Y^{\delta}_{t+k-1}$, and set
$$\begin{aligned} Y_{1,t+k} &= \hat{Q}_{Y_{1,t+k}|Y_{t+k-1}}(\theta_{t+k}) \,, \\ Y^{\delta}_{1,t+k} &= \hat{Q}_{Y_{1,t+k}|Y^{\delta}_{t+k-1}}(\theta_{t+k}) \,. \end{aligned} \tag{12}$$
Note that the same quantile level θ_{t+k} is used to simulate both paths.

3. **Recovering reduced-form shocks on the quantile-augmented variable:** The quantile-based conditional mean forecasts are computed as
$$\begin{aligned} \hat{\tilde{Y}}_{1,t+k} &= \frac{1}{N}\sum_{\tau \in \mathcal{T}_N} \hat{Q}_{Y_{1,t+k}|Y_{t+k-1}}(\tau) \,, \\ \hat{\tilde{Y}}^{\delta}_{1,t+k} &= \frac{1}{N}\sum_{\tau \in \mathcal{T}_N} \hat{Q}_{Y_{1,t+k}|Y^{\delta}_{t+k-1}}(\tau) \,, \end{aligned} \tag{13}$$
and the associated innovations are $\tilde{\epsilon}_{1,t+k} = Y_{1,t+k} - \hat{\tilde{Y}}_{1,t+k}$ and $\tilde{\epsilon}^{\delta}_{1,t+k} = Y^{\delta}_{1,t+k} - \hat{\tilde{Y}}^{\delta}_{1,t+k}$. The structural shocks $\hat{U}_{1,t+k}$ and $\hat{U}^{\delta}_{1,t+k}$ are then recovered using the structural decomposition in Equation (10) and computed as
$$\begin{aligned} \hat{U}_{1,t+k} &= \frac{1}{[\hat{\mathbf{A}}]_{1,1}}\left(\tilde{\epsilon}_{1,t+k} - [\hat{\mathbf{A}}]_{1,-1}\hat{U}_{2,t+k}\right) \\ \hat{U}^{\delta}_{1,t+k} &= \frac{1}{[\hat{\mathbf{A}}]_{1,1}}\left(\tilde{\epsilon}^{\delta}_{1,t+k} - [\hat{\mathbf{A}}]_{1,-1}\hat{U}_{2,t+k}\right) \,, \end{aligned} \tag{14}$$
where $\hat{\mathbf{A}}]_{1,1}$ is the entry in the first row and the first column of matrix $\hat{\mathbf{A}}$, and $[\hat{\mathbf{A}}]_{1,-1}$ is the first row without the first column entry.

4. **Simulating the VAR component:** The structural shock vectors $\hat{U}_{t+k} = (\hat{U}_{1,t+k}, \hat{U}_{2,t+k})'$ and $\hat{U}^{\delta}_{t+k} = (\hat{U}^{\delta}_{1,t+k}, \hat{U}_{2,t+k})'$ then yield the reduced-form shock vectors $\tilde{\epsilon}_{t+k} = \hat{\mathbf{A}}\hat{U}_{t+k}$

more details.

and $\widetilde{\epsilon}^{\delta}_{t+k} = \hat{\mathbf{A}}\hat{U}^{\delta}_{t+k}$.

The values of $Y_{2,t+k}$ and $Y^{\delta}_{2,t+k}$ are obtained as in the VAR by adding the reduced-form innovations to the conditional mean forecasts, yielding

$$\begin{aligned} Y_{2,t+k} &= \hat{\mu}_2 + \hat{\beta}_{2,1} Y_{1,t+k-1} + \hat{\beta}_{2,2} Y^{\delta}_{2,t+k-1} + \hat{\epsilon}_{2,t+k} \\ Y^{\delta}_{2,t+k} &= \hat{\mu}_2 + \hat{\beta}_{2,1} Y^{\delta}_{1,t+k-1} + \hat{\beta}_{2,2} Y^{\delta}_{2,t+k-1} + \hat{\epsilon}^{\delta}_{2,t+k} \ , \end{aligned} \tag{15}$$

or, using a similar notation as above, $Y_{2,t+k} = [\hat{\widetilde{\Lambda}}_{t+k}]_{-1} + [\widetilde{\epsilon}_{t+k}]_{-1}$.

5. **Recursion and repetition:** Repeating steps 2 to 4 recursively for $k = 1, \dots, h$ completes the simulation procedure for a pair of counterfactual paths. Repeat the simulation R times to obtain a sample of trajectories $(Y^{(r)}_{t+k}, Y^{\delta(r)}_{t+k})$ for $k = 1, \dots, h$ and $r = 1, \dots, R$ along with the sorted quantile fitted values $(\hat{Q}^{(r)}_{Y_{t+k}|Y_{t+k-1}}(\tau), \hat{Q}^{\delta(r)}_{Y_{t+k}|Y^{\delta}_{t+k-1}}(\tau))$ for all $\tau \in \mathcal{T}$.

Finally, we estimate the IRF with initial conditions Y_t as the average difference between shocked and baseline simulated paths across all repeated draws

$$\widehat{\text{IRF}}(k, \delta_t) = \frac{1}{R} \sum_{r=1}^{R} Y^{\delta(r)}_{t+k} - Y^{(r)}_{t+k} \ .$$

Similarly, the level-τ QIRF with initial conditions Y_t is estimated as the average difference between fitted quantiles, such that

$$\widehat{\text{QIRF}}(k, \delta_t, \tau) = \frac{1}{R} \sum_{r=1}^{R} \hat{Q}^{\delta(r)}_{Y_{t+k}|Y^{\delta}_{t+k-1}}(\tau) - \hat{Q}^{(r)}_{Y_{t+k}|Y_{t+k-1}}(\tau) \ .$$

Monte Carlo experiments in Appendix C demonstrate that the QAVAR outperforms the linear VAR when the data-generating process features time-varying higher moments. If the true model is linear, the QAVAR slightly biases conditional variance estimates due to the expectation approximation. This trade-off suggests that the QAVAR is most appropriate in settings where higher-order dynamics, such as time-varying skewness or kurtosis, are empirically relevant. Notably, GDP growth is known to exhibit conditional skewness during periods of financial stress (Adrian et al., 2019), a feature we examine further in the following section.

3 Impact of financial shocks on the distribution of GDP

We employ the QAVAR framework to investigate how financial shocks influence the conditional distribution of future GDP growth. While Adrian et al. (2019) document that the distribution of GDP growth exhibits conditional skewness related to financial conditions, their analysis does not isolate the effects of identified financial shocks. In contrast, Gilchrist and Zakrajšek (2012) identify structural financial shocks and examine their impact on GDP growth within a linear VAR, but they do not account for potential asymmetries or higher-moment dynamics in the conditional response. We bring both papers together with our QAVAR.

3.1 Financial conditions shocks in the spirit of Adrian et al. (2019)

We apply our QAVAR to analyze how financial shocks shape the distribution of future GDP growth. Adrian et al. (2019) model the distribution of GDP growth with QR and show that the conditional distribution of GDP growth is skewed due to financial conditions. But they do not identify the impact of financial shocks on the conditional distribution. In this section, we replicate their results with our QAVAR, which also allows for the identification of shocks. In doing so, (i) we can assess the ability of our QAVAR to replicate the results of the ABG model, since both methods focus on the distribution of one single variable, GDP growth, and (ii) we show that their results on financial conditions *changes* carry over to unexpected financial conditions *shocks*.

3.1.1 Model setup

Adrian et al. (2019) use a univariate QR model to consider the impact of the quarterly average of the Chicago Fed's National Financial Conditions Index (NFCI) on real GDP growth four quarters later. They benchmark their results against a bivariate VAR with four lags of real GDP and the NFCI. We run a bivariate QAVAR with four lags against the standard VAR equivalent. Our data starts in 1973Q1 as in the ABG model, until 2019Q4.[10]

We follow the notation of Equation (4) and denote the time-t value of the NFCI as $NFCI_t$, and the log difference of real GDP as GDP_t. As before, the bar notation (i.e., $\overline{NFCI}_t$ and $\overline{GDP}_t$) indicates the conditional mean from a linear VAR; and the tilde is used to denote the quantile-based estimate of the conditional expectation. Specifically, in the present context, we use $\widetilde{GDP}_t$. Furthermore, $\widetilde{\epsilon}_t^{GDP}$ is the quantile-based error term defined

[10] We use the same bivariate setup as Adrian et al. (2019), but we can also add the federal funds rate to bring the model in line with Gilchrist and Zakrajšek (2012). We find broadly similar results under both specifications.

as the difference between the realization and the quantile-based conditional expectation of GDP growth. Thus the QAVAR model is

$$\begin{bmatrix} GDP_t \\ NFCI_t \end{bmatrix} = \begin{bmatrix} \widetilde{GDP_t} \\ \overline{NFCI_t} \end{bmatrix} + \begin{bmatrix} \widetilde{\epsilon}_{GDP,t} \\ \epsilon_{NFCI,t} \end{bmatrix}. \tag{16}$$

Reduced-form residuals $\widetilde{\epsilon}_t = (\widetilde{\epsilon}_{GDP,t}, \epsilon_{NFCI,t})'$ are decomposed into structural shocks $U_t = (U_{GDP,t}, U_{NFCI,t})'$ with $\widetilde{\epsilon}_t = \mathbf{A}U_t$ with the same method as Gilchrist and Zakrajšek (2012), whereby $\mathbf{A}$ is identified with a Cholesky ordering such that real economic shocks affect financial conditions contemporaneously while financial shocks affect the real economy with a lag.

3.1.2 Results

Figure 1 shows the impact of an unexpected financial conditions shock equivalent to a 1 unit increase in the NFCI. In comparison, the quarterly NFCI increased to about 2.6 at the height of the Great Financial Crisis (GFC).

First, we observe that the structural identifications are virtually identical in the QAVAR and the VAR models: on impact, the financial shock has the same effect on the NFCI in both models (top right panel: the solid blue line is identical to the starred black line at period 0). This empirically confirms that the shock identification strategy carries over from the VAR to the QAVAR.

Second, the mean IRF of GDP obtained from the QAVAR (top left panel, solid blue line) is lower than the VAR equivalent (top left panel, starred black line), revealing heterogeneous effects across conditional quantiles of GDP growth. In particular within the first year, the average response of GDP is significantly lower when computed with the QAVAR than with the linear VAR counterpart. This suggests that the estimation of the average effect may be influenced by a wider left-tail of GDP after a financial conditions shock.

Third, the middle and lower panels of Figure 1 show the response of the 5th, 25th, 50th, 75th, and 95th percentiles (as well as the median) of GDP growth in the QAVAR (solid blue lines) against the mean GDP growth response of the VAR (starred black lines). Adrian et al. (2019) use those quantiles to periodically calibrate a skewed-t distribution and generate explicit conditional density estimates of GDP growth. Similar to the ABG model, we find that the 5th percentile (i.e., the 95% Growth-at-Risk) is significantly lower than the mean and other quantiles after a financial conditions shock: the 5th percentile worsens about twice as fast as the mean effect recovered from a linear VAR equivalent.[11] This is also true,

[11]The 10th quantile (not reported) also deteriorates significantly more than the mean, although less than

though to a lesser extent, for the 25th percentile. The impulse response of the expected shortfall (bottom right panel), that is, the average GDP below the 10th percentile, is also significantly below the mean response from the linear VAR.[12] In contrast, quantiles above the median respond homogenously with the mean impulse response. The 95th percentile recovers faster, with a level of GDP not significantly different from its pre-shock value within two years.[13] Altogether, an adverse financial conditions shock generates a negative skew in GDP growth, qualitatively matching the results of Adrian et al. (2019) obtained without structural identification.

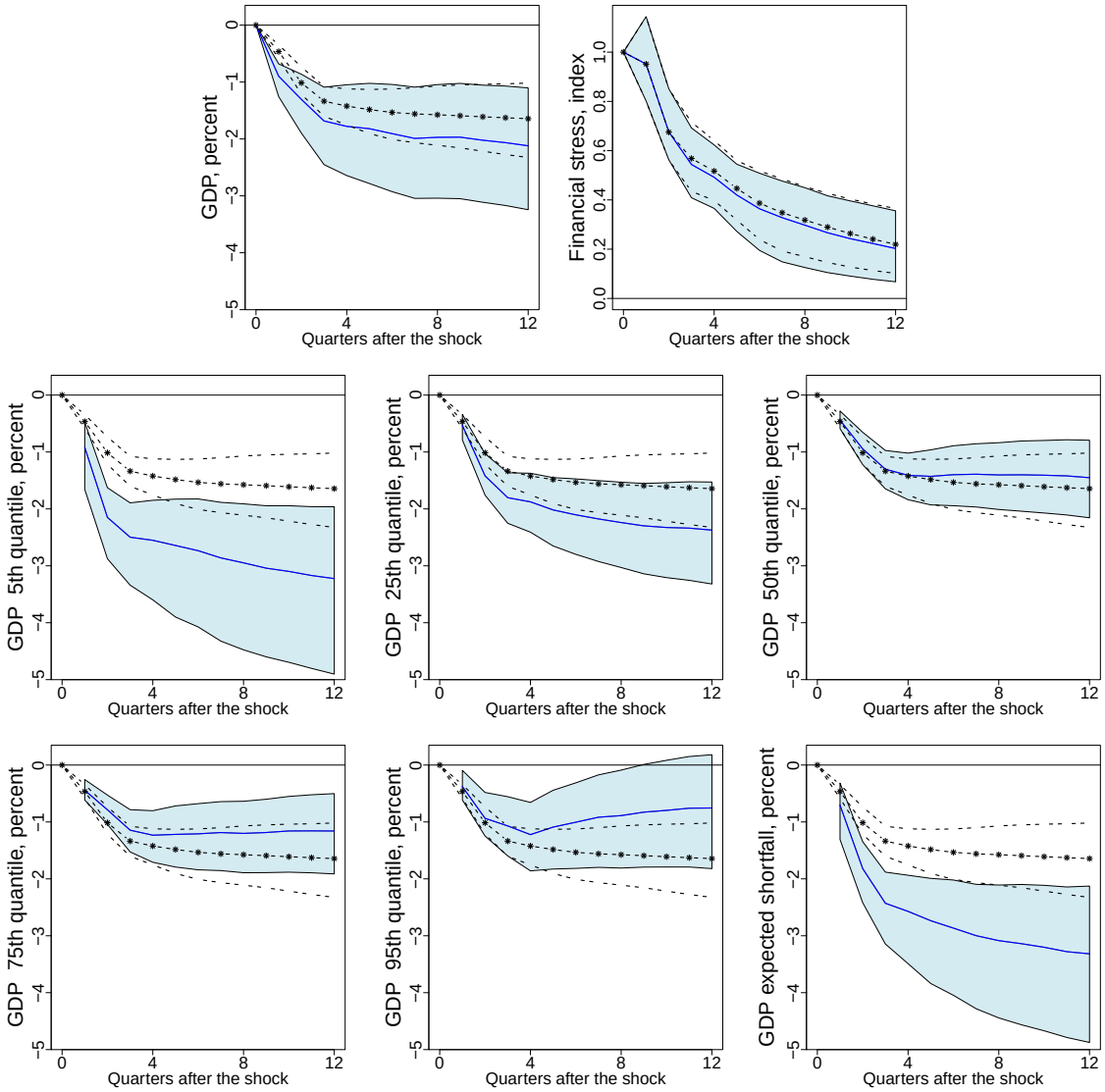
Overall, financial conditions shocks are found to exacerbate downside risk, in line with the literature on Growth-at-Risk models (e.g., Adrian et al., 2019; Aikman et al., 2019; International Monetary Fund, 2017), but also in line with structural VAR with other types of non-nonlinearities (like regime-switching, e.g., Hubrich and Tetlow, 2015, or stochastic volatility, e.g., Carriero et al., 2024), vector moving average representations with non-linear financial terms (Forni et al., 2024), or more structural models of financial frictions (e.g., Adrian et al., 2020; Duprey and Ueberfeldt, 2020).

the 5th percentile. Figure D.1 in Appendix D shows that the lower percentiles are significantly different from the median and the upper percentiles. This implies a Kelly skewness (that theoretically ranges between -1 and 1) dropping by -0.1 in this case.

[12]It is simply computed as the average of the evenly spaced percentiles of GDP growth below the 10th percentile, in simulations with the shock minus the equivalent control without the shock.

[13]Figure D.1 in Appendix D also shows that the upper percentiles are not significantly different from the median, confirming the results of ABG that the upper percentiles of the conditional distribution of future GDP growth are relatively less volatile.

Figure 1: Mean and quantile impulse responses to a financial conditions shock



Notes: The figure shows the response to a financial conditions shock of +1 units of NFCI identified as in Gilchrist and Zakrajšek (2012) and that mimics the univariate results of Adrian et al. (2019), using data from 1973Q1–2019Q4. The VAR model is represented by black starred lines and confidence bands in black dotted lines. The QAVAR model is represented by the solid blue line and the blue shaded confidence bands. The confidence bands are the 10th and 90th percentiles produced with 1000 block-bootstrap replications with block sizes of 4 years and 151 simulated path for each possible initial conditions (151 quarterly observations). The bottom panels show the 5th, 25th, 50th, 75th, and 95th percentile and expected shortfall estimates from the QAVAR (in solid blue) against the mean estimate of the VAR (in dashed black). The response of log-difference of GDP growth (mean and quantiles) is cumulated over the periods.

3.2 Excess bonds premium shocks in the spirit of Gilchrist and Zakrajšek (2012)

We now apply our QAVAR to revisit the study of Gilchrist and Zakrajšek (2012), who document the impact of structural excess bond premium shocks on mean GDP growth in a linear VAR. In doing so, (i) we confirm that the shock identification of the QAVAR is the same as that of the linear VAR, and (ii) we uncover heterogenous effects of excess bond premium shocks not accounted for in the original study.

3.2.1 Model setup

We begin by replicating the baseline results of Gilchrist and Zakrajšek (2012), which estimate the dynamic response of GDP growth to a financial shock captured by innovations in the excess bond premium.[14] The structural VAR contains four lags[15] and three endogenous variables: the log-difference of real output (GDP_t), the quarterly average of the excess bond premium (EBP_t), and the effective nominal federal funds rate (FFR_t), all expressed in percentage points. The sample spans 1973Q1 to 2019Q4 for the United States. The EBP series from 1973Q1 to 2002Q2 is sourced from Gilchrist and Zakrajšek (2012); and from 2002Q3 onward, we use the updated quarterly data from Gilchrist et al. (2021).

Using the notation of Equation (4), we denote the conditional mean of EBP_t and FFR_t as $\overline{EBP}_t$ and $\overline{FFR}_t$. As before, the tilde denotes the quantile-augmented variable, and $\widetilde{GDP}_t$ is hence the quantile-based approximation of the conditional expectation of GDP_t and $\widetilde{\epsilon}_{GDP,t} = GDP_t - \widetilde{GDP}_t$ is the associated error term. The QAVAR counterpart to the GZ model then takes the form

$$\begin{bmatrix} GDP_t \\ EBP_t \\ FFR_t \end{bmatrix} = \begin{bmatrix} \widetilde{GDP}_t \\ \overline{EBP}_t \\ \overline{FFR}_t \end{bmatrix} + \begin{bmatrix} \widetilde{\epsilon}_{GDP,t} \\ \epsilon_{EBP,t} \\ \epsilon_{FFR,t} \end{bmatrix}, \tag{17}$$

and the residuals $\widetilde{\epsilon}_t = (\widetilde{\epsilon}_{GDP,t}, \epsilon_{EBP,t}, \epsilon_{FFR,t})'$ can be decomposed into structural shocks $U_t = (U_{GDP,t}, U_{EBP,t}, U_{FFR,t})'$, with $\widetilde{\epsilon}_t = \mathbf{A}U_t$ and $\mathbf{A}$ identified with a Cholesky decomposition. The identification strategy for financial shocks relies on the recursive ordering of the variables,

[14] Note that the GZ model includes, in order, consumption, investment, output growth, prices, the excess bond premium, the cumulated excess market return, the 10-year Treasury yield, and the federal funds rate. Using the authors' code, we verified that the dynamics of output growth, the excess bond premium, and the federal funds rate are very similar when using only the three variables with the same ordering. We focus on this trivariate version for simplicity.

[15] GZ's original model uses two lags, but we confirmed that four lags generate similar results. Using four lags in our QAVAR brings more consistency with the ABG setup discussed in the previous section and the GDP growth-at-risk peaking in the first four quarters (Adrian et al., 2022; Duprey and Ueberfeldt, 2020).

with shocks to the excess bond premium affecting the real economy with a lag, while the risk-free rate can react contemporaneously to the financial disturbance.

3.2.2 Results

Figure 2 shows the impact of an unexpected financial shock equivalent to a 1 percentage point increase in the EBP. In comparison, the EBP increased to about 4 percentage points around the GFC and about 2 percentage points during the pandemic (c.f., Figure 2, Gilchrist et al. (2021)).

First, we observe that structural shocks identified in the GZ model again match those of the QAVAR model. On impact, the financial shock has the same effect on the EBP in both models (top middle panel: the solid blue line is identical to the starred black line at period 0). This empirically confirms that the shock identification strategy carries over from the VAR to the QAVAR.

Second, the mean response of cumulative GDP obtained from the QAVAR (top left panel, solid blue line) is broadly similar to the VAR equivalent (top left panel, starred black line). However, the response of various conditional quantiles of GDP may be different if the mean estimator hides state-contingent heterogeneity.

Third, the middle and lower panels of Figure 2 show the response of the 5th, 25th, 50th, 75th, and 95th percentiles of GDP growth in the QAVAR (solid blue lines) against the mean GDP response of the VAR (starred black lines). The response of the 5th quantile (left middle panel) is significantly different from zero in the first two years and initially significantly worse than the mean VAR estimate (the response of the quantile, the blue line, is outside the confidence band of the mean VAR response, the dashed black lines).[16] The response of the 5th percentile is about twice as negative as the mean response in the linear VAR. Similarly, the expected shortfall that occurs with a 10% probability is largely negative outside the confidence bands of the linear VAR. For the median (as well as the upper quantiles), the correction of GDP is significantly more muted than the mean VAR estimator. Thus, overall, an excess bonds premium shock also tends to be a skewness shock.

Although financial shocks associated with larger EBP are detrimental to the economy, they are less detrimental to the median GDP growth and upside risks, and more detrimental to downside risks, just like in the case of a financial conditions shock and the related literature mentioned in the previous section. Our results are also consistent with Loria et al. (2024), who use the changes in the excess bond premium in a local projection model esti-

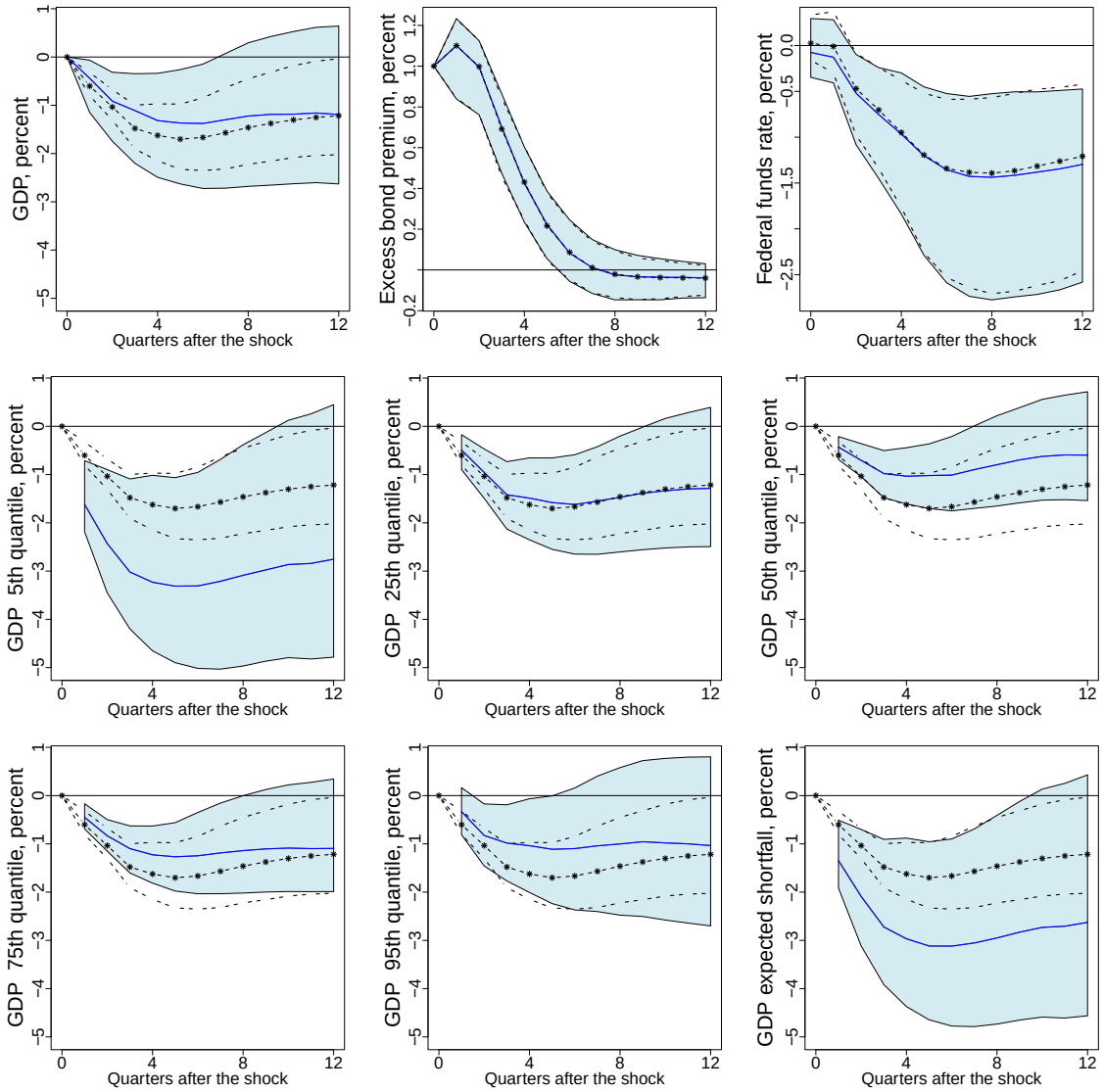
[16]The response of the lower percentile is also significantly different from the median and the upper percentile, as seen Figure D.2 in Appendix D. This implies a Kelly skewness (that theoretically ranges between -1 and 1) dropping by -0.15.

mated with QR. Within a year, the 10th percentile of GDP growth decreases more than the median or the 90th percentile. Compared to the quantile local projection approach that requires exogenously identified shocks, our method allows for the shock identification, dynamic evolution, and tail elasticities within the same framework.

We note that although financial conditions shocks and excess bond premium shocks both worsen downside risks to GDP growth, excess bond premium shocks have a relatively less significant impact on the tail than financial conditions shocks.[17] Maybe this is not too surprising given that the NFCI is a broader composite of various types of financial markets. This means that a worsening of the NFCI has a systemic aspect, occurring only during severe financial crises, while excess bond premium shocks are focused only on the corporate bonds market segment.

[17]Note that this not driven by the choice of a bivariate VAR to match ABG's setup versus a trivariate VAR to match GZ's setup: adding the federal funds rate to Equation (16) generates similar results.

Figure 2: Mean and quantile impulse responses to an excess bonds premium shock



Notes: The figure shows the response to a financial shock of +1 percentage points of EBP taken from Gilchrist and Zakrajšek (2012) for 1973Q1–2002Q2 and from Gilchrist et al. (2021) for 2002Q3–2019Q4. The VAR model is represented by black starred lines and confidence bands by black dotted lines. The QAVAR model is represented by the solid blue line and the blue shaded confidence bands. The confidence bands are the 10th and 90th percentiles produced with 1000 block-bootstrapped replications with block sizes of 4 years and 151 simulations for each possible initial conditions (151 quarterly observations). The mean response in the VAR shown in the top panels are similar to the Figure 5 of Gilchrist and Zakrajšek (2012). The bottom panels show the 5th, 25th, 50th, 75th, and 95th percentile and expected shortfall estimates from the QAVAR (in solid blue) against the mean estimate of the VAR (in dashed black). The response of log-difference of GDP growth (mean and quantiles) is cumulated over the periods.

4 Conclusion

We study how financial shocks affect not only average GDP growth but also the full distribution of possible outcomes—from deep recessions to strong recoveries. Using a simple extension of the standard VAR framework that incorporates regressed conditional quantiles—a quantile-augmented VAR (QAVAR)—we explore the effect of structural financial shocks on the distribution of future GDP growth.

To highlight the importance of assessing risks across the entire distribution, we revisit the influential modelling frameworks of Adrian et al. (2019) and Gilchrist and Zakrajšek (2012). First, we show that financial conditions *shocks* identified structurally significantly worsen growth-at-risk, replicating the core insights of Adrian et al. (2019) while adding structural identification. Second, we demonstrate that excess bond premium shocks also alter the shape of the distribution of GDP, an effect averaged out in the linear VAR of Gilchrist and Zakrajšek (2012). These shocks are less harmful to the median and upper percentiles of the GDP growth distribution, but they deepen the risk of downturns.

Our framework offers policymakers a tractable tool to better assess the risks financial shocks pose to economic growth. By preserving standard shock identification while allowing for nonlinear dynamics, the QAVAR enhances our understanding of macro-financial linkages and the distributional consequences of financial stress. For example, in a world of heightened tariff-related uncertainty and financial fragility, our findings suggest that corporate bond market stress may serve as an early warning of a deeper recession than previous estimates would otherwise suggest.

Future extensions of the QAVAR could explore the role of uncertainty shocks on the interquantile range of GDP growth, for instance by incorporating uncertainty measures directly into the VAR or decomposing quantile-specific shocks. Additional research could also extend the framework to jointly model the distribution of multiple variables or explore how to identify structural shocks that target specific parts of the distribution.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable Growth," *American Economic Review*, 109, 1263–89. 1, 2.5, 2.5, 3, 3.1, 3.1.1, 10, 3.1.2, 1, 4, 3, C.3, D.1

Adrian, T., F. Duarte, N. Liang, and P. Zabczyk (2020): "NKV: A New Keynesian Model with Vulnerability," *AEA Papers and Proceedings*, 110, 470–76. 3.1.2

Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022): "The Term Structure of Growth-at-Risk," *American Economic Journal: Macroeconomics*, 14, 283–323. 15, C.3

Aikman, D., J. Bridges, S. Hacioglu Hoke, C. O'Neill, and A. Raja (2019): "Credit, Capital and Crises: A GDP-at-Risk approach," Bank of England working papers 824, Bank of England. 3.1.2

Arisoy, Y. E. (2023): "Quantile Regression: Estimating Moments of the Stock Return Distribution," *Available at SSRN 4515136*. 1

Beaudry, P. and G. Koop (1993): "Do Recessions Permanently Change Output?" *Journal of Monetary Economics*, 31, 149–163. 6

Bellemare, M. G., W. Dabney, and R. Munos (2017): "A Distributional Perspective on Reinforcement Learning," in *International Conference on Machine Learning*, PMLR, 449–458. 1

Bondell, H. D., B. J. Reich, and H. Wang (2010): "Noncrossing Quantile Regression Curve Estimation," *Biometrika*, 97, 825–838. 4

Cao, J., J. Chen, J. Hull, and Z. Poulos (2020): "Deep Hedging of Derivatives Using Reinforcement Learning," *The Journal of Financial Data Science*. 1

Carriero, A., T. E. Clark, and M. Marcellino (2024): "Capturing Macro-Economic Tail Risks with Bayesian Vector Autoregressions," *Journal of Money, Credit and Banking*, 56, 1099–1127. 3.1.2

Chavleishvili, S., R. F. Engle, S. Fahr, M. Kremer, S. Manganelli, and B. Schwaab (2021): "The Risk Management Approach to Macro-Prudential Policy," Working Paper Series 2565, European Central Bank. 1

Chavleishvili, S. and S. Manganelli (2024): "Forecasting and Stress Testing with Quantile Vector Autoregression," *Journal of Applied Econometrics*, 39, 66–85. 1, 3

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125. 4

Dabney, W., M. Rowland, M. Bellemare, and R. Munos (2018): "Distributional Reinforcement Learning with Quantile Regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32. 1

Davino, C., M. Furno, and D. Vistocco (2013): *Quantile Regression: Theory and Applications*, vol. 988, John Wiley & Sons. 2.5

Davis, C. (2015): "The Skewed Generalized T Distribution Tree Package Vignette," University of Chicago. 3

Duprey, T. and A. Ueberfeldt (2020): "Managing GDP Tail Risk," Staff working paper, 2020-03, Bank of Canada. 3.1.2, 15

Engle, R. F. and S. Manganelli (2004): "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, 367–381. 1

Forni, M., L. Gambetti, N. Maffei-Faccioli, and L. Sala (2024): "Nonlinear Transmission of Financial Shocks: Some New Evidence," *Journal of Money, Credit and Banking*, 56, 5–33. 3.1.2

Gaglianone, W. P. and L. R. Lima (2012): "Constructing Density Forecasts from Quantile Regressions," *Journal of Money, Credit and Banking*, 44, 1589–1607. 2.5

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993): "Nonlinear Dynamic Structures," *Econometrica: Journal of the Econometric Society*, 871–907. 2.5

Gilchrist, S., B. Wei, V. Z. Yue, and E. Zakrajšek (2021): "The Term Structure of the Excess Bond Premium: Measures and Implications," Policy Hub 2021-12, Federal Reserve Bank of Atlanta. 3.2.1, 3.2.2, 2, D.2

Gilchrist, S. and E. Zakrajšek (2012): "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 102, 1692–1720. 1, 3, 10, 3.1.1, 1, 3.2, 3.2.1, 2, 4, D.1, D.2

He, X. (1997): "Quantile Curves Without Crossing," *The American Statistician*, 51, 186–192. 4

Hubrich, K. and R. J. Tetlow (2015): "Financial Stress and Economic Dynamics: The Transmission of Crises," *Journal of Monetary Economics*, 70, 100–115. 3.1.2

International Monetary Fund (2017): "Financial Conditions and Growth at Risk," Chapter 3 of the *Global Financial Stability Report: Is Growth at Risk?* 3.1.2

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015): "Generalised density forecast combinations," *Journal of Econometrics*, 188, 150–165. 8

Koenker, R. (2005): *Quantile Regression*, Econometric Society Monographs, Cambridge University Press. 2.5

Koenker, R. and G. Bassett Jr (1978): "Regression Quantiles," *Econometrica: Journal of the Econometric Society*, 33–50. 2.2, 2.3

Koop, G., M. H. Pesaran, and S. M. Potter (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147. 2.5

Korobilis, D. (2017): "Quantile Regression Forecasts of Inflation Under Model Uncertainty," *International Journal of Forecasting*, 33, 11–20. 2.5

Loria, F., C. Matthes, and D. Zhang (2024): "Assessing Macroeconomic Tail Risk," *The Economic Journal*, 135, 264–284. 1, 3.2.2

Potter, S. M. (2000): "Nonlinear impulse response functions," *Journal of Economic Dynamics and Control*, 24, 1425–1446. 2.5

Reinhart, C. M. and K. S. Rogoff (2014): "Recovery from Financial Crises: Evidence from 100 Episodes," *American Economic Review*, 104, 50–55. 1, 18

Schmidt, L. and Y. Zhu (2016): "Quantile Spacings: A Simple Method for the Joint Estimation of Multiple Quantiles Without Crossing," *Available at SSRN 2220901*. 4, 2.5

White, H., T.-H. Kim, and S. Manganelli (2015): "VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles," *Journal of Econometrics*, 187, 169–188. 1

A Analytical impulse response function of the conditional mean

In the structural VAR given by Equation (1), the conditional mean response to a one-time shock $\delta_t = (\delta_{1,t}, \delta_{2,t})'$ at time t admits a closed-form solution. Consider a sequence of structural innovations, $\{U_t, \ldots, U_{t+k}\}$, over a horizon of length, $k > 0$. The resulting benchmark path is recursively given by

$$(Y_{t+k} - \mu) = \mathbf{B}^k (Y_t - \mu) + \sum_{j=0}^{k} \mathbf{B}^j \mathbf{A} U_{t+k-j} \ . \tag{18}$$

Under the counterfactual scenario, we assume the shock at time t is perturbed by δ_t, yielding the modified sequence $\{U_t + \delta_t, U_{t+1}, \ldots, U_{t+k}\}$. The shocked path then satisfies

$$(Y^{\delta}_{t+k} - \mu) = \mathbf{B}^k (Y^{\delta}_t - \mu) + \sum_{j=0}^{k} \mathbf{B}^j \mathbf{A} U_{t+k-j} \ , \tag{19}$$

where $Y^{\delta}_t = Y_t + \mathbf{A}\delta_t$ is the contemporaneous shocked value at time t. Taking expectations of the difference between the shocked and baseline paths conditional on Y_{t-1}, we obtain the IRF as

$$\text{IRF}(k, \delta_t) = \mathbb{E}_t[Y^{\delta}_{t+k} - Y_{t+k}] = \mathbf{B}^k \mathbf{A} \delta_t \ . \tag{20}$$

An analogous recursion characterizes the dynamics of the QAVAR system defined in Equation (9). Given a sequence of structural innovations $\{\hat{U}_t, \ldots, \hat{U}_{t+k}\}$, the baseline path evolves as

$$(Y_{t+k} - \hat{\tilde{\mu}}) = \hat{\tilde{\mathbf{B}}}^k (Y_t - \hat{\tilde{\mu}}) + \sum_{j=0}^{k} \hat{\tilde{\mathbf{B}}}^j \hat{\mathbf{A}} \hat{U}_{t+k-j} \ . \tag{21}$$

Following the same logic, we define a counterfactual scenario where the initial shock is perturbed: $\hat{U}^{\delta}_t = \hat{U}_t + \delta_t$. Unlike the linear case, however, all subsequent shocks $\{\hat{U}^{\delta}_{t+1}, \ldots, \hat{U}^{\delta}_{t+k}\}$ are drawn from distributions induced by the shocked path. The resulting trajectory is given by

$$(Y^{\delta}_{t+k} - \hat{\tilde{\mu}}) = \hat{\tilde{\mathbf{B}}}^k (Y^{\delta}_t - \hat{\tilde{\mu}}) + \sum_{j=0}^{k} \hat{\tilde{\mathbf{B}}}^j \hat{\mathbf{A}} \hat{U}^{\delta}_{t+k-j} \ . \tag{22}$$

Assuming that baseline and structural shocks $\hat{U}_t$ and $\hat{U}^{\delta}_t$ both have a mean approaching zero, we approximate the estimated impulse response as

$$\widehat{\text{IRF}}(k, \delta_t) \approx \hat{\tilde{\mathbf{B}}}^k \hat{\mathbf{A}} \delta_t \ . \tag{23}$$

This expression characterizes the impact of the structural shock δ_t on the conditional mean within the QAVAR framework.

Importantly, this IRF remains an *approximation* whose accuracy depends on the granularity of the quantile grid $\mathcal{T}_N$ used to estimate the underlying conditional distribution.

B Bootstrapped confidence intervals

In this section, we describe the bootstrap procedure used to construct confidence intervals for the mean and quantile impulse response functions (IRFs and QIRFs, respectively) reported in Section 3. We adopt a block bootstrap method to resample structural residuals and generate a distribution of simulated impulse responses, from which we derive confidence bands.[18]

Starting from the estimated QAVAR model in Equation(9), we extract the full set of reduced-form residuals $\widehat{\widetilde{\epsilon}}_t$. We then generate a synthetic residual series of the same length by resampling blocks of size S, resulting in a sequence, $\{\widetilde{\epsilon}^*_t = (\widetilde{\epsilon}^{(b)}_{1,t}, \epsilon^{(b)}_{2,t})' : t = 1, \ldots, T\}$, where the superscript (b) denotes bootstrapped quantities. The use of block sampling captures potential remaining serial dependence of the estimated structural errors.

The synthetic sample $\{Y^{(b)}_t : t = 1, \ldots, T\}$ is constructed recursively as follows:

1. **Initialization:** Select an initial value $Y^{(b)}_0$. Draw $\lceil T/S \rceil$ blocks of size S from $\{\widehat{\widetilde{\epsilon}}_t : t = 1, \ldots, T\}$ using simple random sampling, and concatenate them to form the synthetic series $\{\widetilde{\epsilon}^{(b)}_t = (\widetilde{\epsilon}^{(b)}_{1,t}, \epsilon^{(b)}_{2,t})' : t = 1, \ldots, T\}$.

2. **Mean forecast:** For $t = 1, \ldots, T$, use the original QAVAR estimates $\hat{\mu}$ and $\hat{\mathbf{B}}$ to construct the mean forecast, defined as $\widetilde{\Lambda}^{(b)}_t = \mathbf{W}(\hat{\mu} + \hat{\mathbf{B}} Y^{(b)}_{t-1})$.

3. **Realized value:** Compute the synthetic outcome as the sum of the forecast and bootstrapped innovation, such that $Y^{(b)}_t = \widetilde{\Lambda}^{(b)}_t + \widetilde{\epsilon}^*_t$.

4. **Recursion:** Repeat steps 2 and 3 iteratively for all $k = 1, \ldots, T$ to generate a complete bootstrapped sample $\{Y^{(b)}_t : t = 1, \ldots, T\}$.

5. **Re-estimation and identification:** Estimate the QAVAR model using the synthetic data to obtain $\hat{\mu}^{(b)}$, $\hat{\mathbf{B}}^{(b)}$, and the associated fitted values $\hat{\widetilde{\Lambda}}^{(b)}_t = \mathbf{W}(\hat{\mu}^{(b)} + \hat{\mathbf{B}}^{(b)} Y^{(b)}_{t-1})$ as well as the corresponding estimated residuals $\hat{\widetilde{\epsilon}}^{(b)}_t$ for $t = 1, \ldots, T$. These are then

[18] The median and mean peak-to-trough durations for 63 crises in advanced economies are 2 and 2.9 years, respectively (Reinhart and Rogoff, 2014). This supports the use of block bootstrapping with blocks longer than two years to appropriately capture the dynamics of entire crisis episodes.

used in the structural identification procedure from Section 2.4 to compute $\hat{\mathbf{A}}^{(b)}$, the bootstrapped analogue to $\hat{\mathbf{A}}^{(N)}$ in Equation (10).

6. **IRF and QIRF simulation:** For a given structural shock δ_t and a horizon h, simulate R pairs of counterfactual paths using the algorithm outlined in Section 2.5 and compute the IRF and QIRF estimates.

Steps 1 to 6 are repeated independently to generate a sample of B independent bootstrapped IRFs, denoted $\{\widehat{\text{IRF}}^{(b)}(k, \delta_t)\}$ for $k = 1, \ldots, h$ and $b = 1, \ldots, B$, along with a sample of level-τ bootstrapped QIRFs, denoted $\{\widehat{\text{QIRF}}^{(b)}(k, \delta_t, \tau)\}$ for all $\tau \in \mathcal{T}_N$. Finally, for each horizon $k = 1, \ldots, h$, we construct confidence bands based on empirical quantiles of the sets $\{\widehat{\text{IRF}}^{(b)}(k, \delta_t)\}$ and $\{\widehat{\text{QIRF}}^{(b)}(k, \delta_t, \tau)\}$.

C Monte Carlo study

This appendix evaluates the finite-sample forecasting performance of the QAVAR relative to a standard Gaussian VAR. The analysis is conducted via Monte Carlo simulations under various data-generating processes (DGPs) that differ in terms of conditional heteroskedasticity and higher-order moment dynamics. The QAVAR is designed to approximate the full conditional distribution of a target variable, and its performance is assessed in forecasting conditional mean, variance, and skewness. Results indicate that while both models yield comparable accuracy in conditional mean and variance forecasts, the QAVAR provides a substantial improvement in characterizing skewed distributions.

C.1 Data generating processes

We simulate data from a class of bivariate autoregressive processes with time-varying volatility and, in some cases, asymmetric innovations. Let $X_t = (X_{1,t}, X_{2,t})'$ denote the bivariate process at time t, evolving according to

$$\begin{aligned} X_t &= \Phi X_{t-1} + \upsilon_t \\ \upsilon_t &= \sigma_t \xi_t \\ \sigma_t^2 &= \omega + \Gamma \upsilon_{t-1}^2 \ , \end{aligned} \tag{24}$$

where $\xi_t = (\xi_{1,t}, \xi_{2,t})'$ is a standard i.i.d. innovation process, and $\sigma_t^2 = \mathbb{E}[\upsilon_t \upsilon_t'] = \text{diag}(\sigma_{1,t}^2, \sigma_{2,t}^2)$ is a diagonal variance-covariance matrix. The matrices Φ, ω, and Γ are specified to ensure covariance-stationarity of X_t. We consider three DGP specifications:

1. Gaussian AR(1):

$$\Gamma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \xi_t \sim N(\mathbf{0}, I_2) \ ;$$

2. Gaussian ARCH(1,1):

$$\Gamma = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \text{ and } \xi_t \sim N(\mathbf{0}, I_2) \ ;$$

3. Skewed ARCH(1,1):

$$\Gamma = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \text{ and } \xi_{1,t} \sim ST(0, 1, \lambda{=}0.25), \quad \xi_{2,t} \sim N(0, 1) \ ,$$

   where $ST(\lambda)$ denotes the standardized skewed-t distribution with slant parameter $\lambda \in (-1, 1)$ as in Davis (2015). See Adrian et al. (2019) for further discussion on the relevance of the skewed-t distribution in macroeconomic data.

In all specifications, the remaining parameter matrices are set to

$$\omega = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \text{ and } \quad \Phi = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{bmatrix} .$$

C.2 Simulation results

For each DGP, we generate 1000 Monte Carlo replications with sample sizes $T \in \{200, 500, 1000\}$. We estimate one-lag VAR and QAVAR models with $N = 99$ equidistant quantile levels $\mathcal{T}_N$ and extract the first three conditional moments for the target variable $X_{1,t}$. While the Gaussian VAR delivers these moments analytically, the QAVAR relies on numerical integration. Specifically, the quantile-based approximation of the m-th conditional moment $\mathbb{E}_{t-1}[Y_{1,t}^m]$ is such that

$$\frac{1}{N} \sum_{\tau \in \mathcal{T}_N} (Q_{Y_{1,t}|Y_{t-1}}(\tau))^m \to \int_0^1 (Q_{Y_{1,t}|Y_{t-1}}(\tau))^m d\tau = \mathbb{E}_{t-1}[Y_{1,t}^m]$$

as $N \to \infty$.

Table C.1 reports the root mean squared error (RMSE) of one-step-ahead forecasts for the conditional mean, variance, and skewness. Under Gaussian innovations (i.e., the first two DGPs), both models perform comparably in forecasting the conditional mean, and the standard VAR slightly outperforms the QAVAR in variance forecasts when volatility is constant. However, in the skewed DGP, the QAVAR demonstrates a substantial improvement

in forecasting conditional skewness, with RMSE reduction factors of up to about 3 compared to the VAR.

Table C.1: Mean RMSE of conditional mean, variance, and skewness one-step-ahead forecasts

	Gaussian AR(1)			Gaussian ARCH(1,1)			Skewed ARCH(1,1)		
Sample size	200	500	1000	200	500	1000	200	500	1000
RMSE of the conditional mean									
VAR	**.078**	**.051**	**.036**	.124	.078	.057	.122	**.077**	**.055**
	(.032)	(.022)	(.015)	(.053)	(.033)	(.023)	(.053)	(.032)	(.024)
QAVAR	.079	**.051**	**.036**	**.122**	**.077**	**.056**	**.120**	**.077**	.057
	(.033)	(.022)	(.015)	(.052)	(.032)	(.022)	(.051)	(.031)	(.023)
RMSE of the conditional variance									
VAR	**.040**	**.025**	**.018**	**.506**	**.503**	**.502**	**.514**	**.509**	**.508**
	(.030)	(.019)	(.013)	(.050)	(.031)	(.024)	(.056)	(.035)	(.025)
QAVAR	.093	.064	.051	.519	.512	.509	.524	.513	.511
	(.043)	(.028)	(.022)	(.059)	(.035)	(.026)	(.062)	(.038)	(.027)
RMSE of the conditional skewness									
VAR	**0**	**0**	**0**	**0**	**0**	**0**	.374	.373	.383
	(0)	(0)	(0)	(0)	(0)	(0)	(.169)	(.109)	(.076)
QAVAR	.289	.190	.134	.304	.205	.147	**.277**	**.189**	**.141**
	(.118)	(.078)	(.056)	(.129)	(.084)	(.061)	(.119)	(.075)	(.057)

Notes: The table shows the mean RMSE for the fitted conditional mean, variance, and skewness on variable X_1 governed by the three specifications of the DGP of Equation (24) for VAR and QAVAR models. RMSE standard errors are reported in parentheses for 1000 repeated simulations with 200, 500, and 1000 observations. In each case, the lowest mean RMSE is printed in boldface.

Table C.2 complements this analysis by displaying the forecast bias for the same conditional moments. Across all DGPs, the conditional mean forecasts are effectively unbiased under both models. The QAVAR exhibits a minor positive bias in variance forecasts, particularly when volatility is constant. Most importantly, in the presence of skewed innovations, the QAVAR achieves a skewness bias reduction factor of about 10, with negligible distortion in lower-order moments.

C.3 Sensitivity to distributional asymmetry

Figures C.1 and C.2 focus on the skewed ARCH(1,1) and evaluate the robustness of the QAVAR for varying degrees of skewness of the DGP. Figure C.1 displays mean RMSEs of the conditional mean, variance and skewness one-step-ahead forecasts across a range of slant parameter values $\lambda \in \{-0.9, -0.8, \ldots, -0.1, 0, 0.1, \ldots, 0.8, 0.9\}$. Each plot aggregates results

Table C.2: Bias of conditional mean, variance, and skewness one-step-ahead forecasts

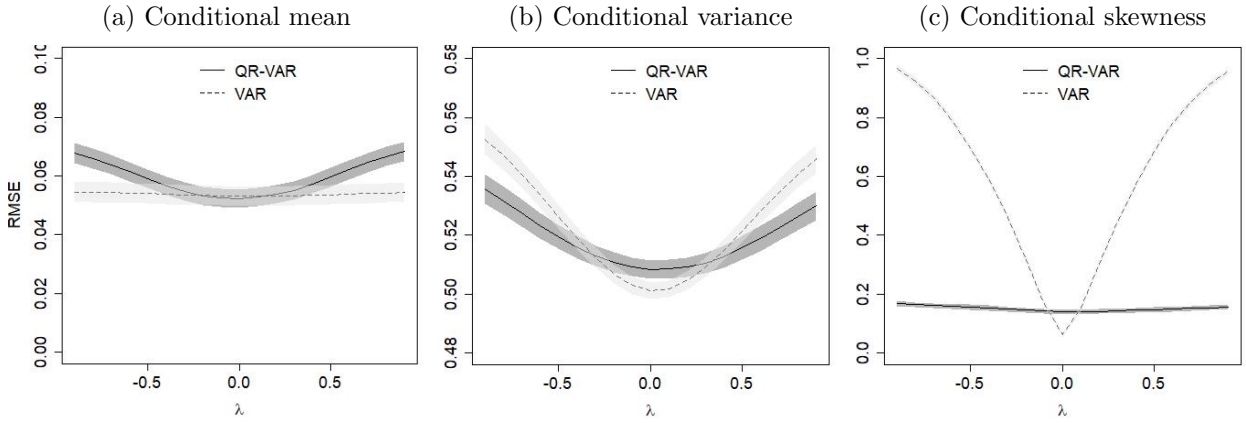
	Gaussian AR(1)			Gaussian ARCH(1,1)			Skewed ARCH(1,1)		
Sample size	200	500	1000	200	500	1000	200	500	1000
Bias of the conditional mean									
VAR	**.000**	**.001**	**.000**	**.001**	**-.004**	**.000**	**.000**	**-.002**	**-.002**
	(.048)	(.033)	(.023)	(.069)	(.044)	(.031)	(.069)	(.045)	(.032)
QAVAR	**.000**	**.001**	**.000**	**.001**	**-.004**	**.000**	-.008	-.009	-.009
	(.048)	(.033)	(.023)	(.067)	(.043)	(.031)	(.067)	(.044)	(.031)
Bias of the conditional variance									
VAR	**-.004**	**-.002**	**-.001**	**-.012**	**-.004**	**-.003**	**-.014**	**-.006**	**-.002**
	(.049)	(.032)	(.022)	(.108)	(.073)	(.051)	(.114)	(.073)	(.053)
QAVAR	.038	.035	.035	.033	.032	.031	.027	.026	.028
	(.054)	(.035)	(.024)	(.104)	(.068)	(.048)	(.103)	(.066)	(.048)
Bias of the conditional skewness									
VAR	**0**	**0**	**0**	**0**	**0**	**0**	-.373	-.373	-.383
	(0)	(0)	(0)	(0)	(0)	(0)	(.171)	(.109)	(.076)
QAVAR	-.006	.001	.000	-.007	-.002	-.002	**-.036**	**-.038**	**-.035**
	(.182)	(.122)	(.083)	(.206)	(.132)	(.096)	(.152)	(.093)	(.071)

Notes: The table reports the mean forecast bias for the conditional mean, variance, and skewness of variable X_1 under the three specifications of the DGP in Equation (24). Biases are averaged over 1000 Monte Carlo replications with sample sizes of 200, 500, and 1000 observations. Standard errors are reported in parentheses. The lowest absolute bias is shown in boldface.

over 200 repeated simulations with $T = 1000$ observations. The QAVAR and VAR models yield indistinguishable forecasting performance for the mean and variance (Panels (a) and (b)), regardless of the level of asymmetry. However, as seen in Panel (c), the QAVAR delivers markedly improved skewness forecasts as $|\lambda|$ approaches 1, confirming its superior ability to capture asymmetry.

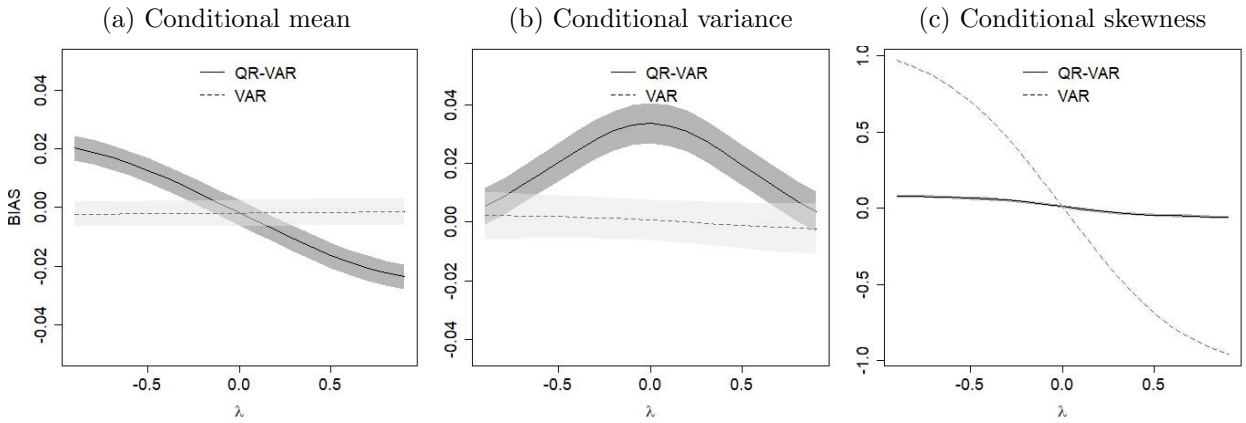
Altogether, these results indicate that the QAVAR is well-suited for environments in which higher-order conditional moments are of interest, and skewness in particular. Indeed, while its performance in forecasting the conditional mean and variance remains comparable to that of a standard VAR, its ability to approximate conditional skewness is significantly superior in non-Gaussian settings. This makes the QAVAR a compelling alternative for modelling macroeconomic variables known to exhibit asymmetric dynamics, such as GDP growth (Adrian et al., 2019, 2022).

Figure C.1: Mean RMSE of the conditional mean, variance, and skewness one-step-ahead forecasts for different slant parameter of the skewed-t distribution



Notes: The panels show the mean RMSE for the fitted conditional mean, variance, and skewness forecast on variable X_1 in the skewed ARCH(1,1) specification of the DGP of Equation (24) for VAR and QAVAR models, along with 95% confidence intervals. Results are computed with 200 repeated simulations for slant paramaters $|\lambda| \in \{0, 0.1, \ldots, 0.9\}$.

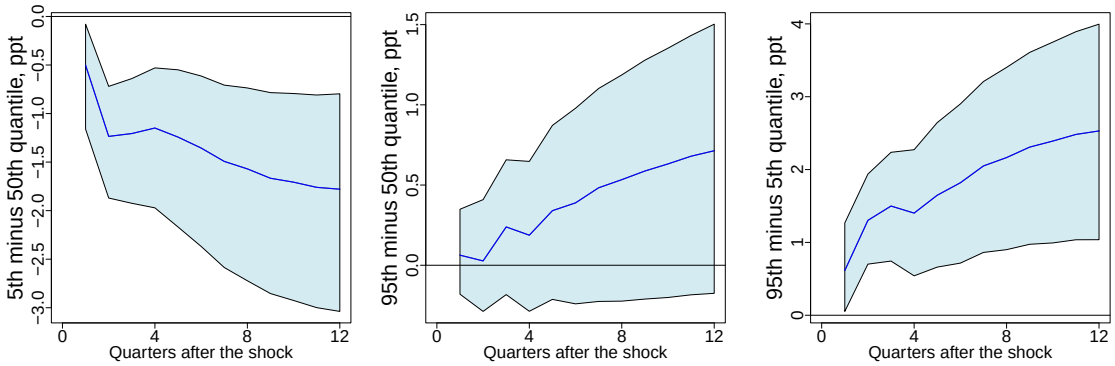
Figure C.2: Bias of the conditional mean, variance, and skewness one-step-ahead forecasts for different slant parameter of the skewed-t distribution



Notes: The panels show the bias for the conditional mean, variance, and skewness forecast on variable X_1 in the skewed ARCH(1,1) specification of the DGP of Equation (24) for VAR and QAVAR models, along with 95% confidence intervals. Results are computed with 200 repeated simulations for slant paramaters $|\lambda| \in \{0, 0.1, \ldots, 0.9\}$.

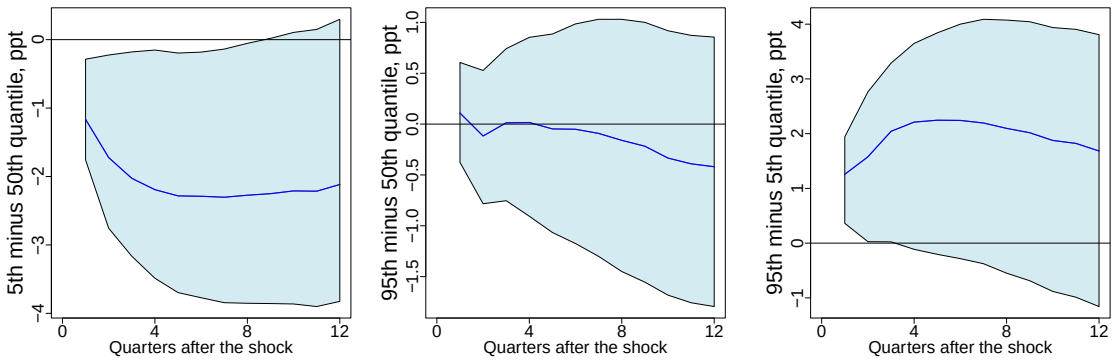
D Additional empirical results

Figure D.1: Difference across quantile responses after a financial conditions shock



Notes: The figure shows the response to a financial conditions shock of +1 unit of NFCI identified as in Gilchrist and Zakrajšek (2012) and that mimics the univariate results of Adrian et al. (2019), using data from 1973Q1–2019Q4. Difference across quantiles from the QAVAR model. The confidence bands are the 10th and 90th percentiles produced with 1000 block-bootstrap replications with block sizes of four years and 151 simulated path for each possible initial conditions (151 quarterly observations). The response of log-difference of GDP growth is cumulated over the periods.

Figure D.2: Difference across quantile responses after an excess bonds premium shock



Notes: The figure shows the response to a financial shock of +1 percentage point of EBP taken from Gilchrist and Zakrajšek (2012) for 1973Q1–2002Q2 and from Gilchrist et al. (2021) for 2002Q3–2019Q4. Difference across quantiles from the QAVAR model. The confidence bands are the 10th and 90th percentiles produced with 1000 block-bootstrapped replications with block sizes of four years and 151 simulations for each possible initial conditions (151 quarterly observations). The response of log-difference of GDP growth is cumulated over the periods.