Understanding the State-Dependent Impact of Financial Shocks on Growth via Unconditional Quantile Impulse Responses.

Christian Brownlees, Robert Wojciechowski

June 26, 2023

Abstract

A large and growing body of empirical literature seeks to estimate state-dependent impulse responses of output growth to financial shocks. However, state-dependence when the state is a function of the outcome variable is particularly problematic in the conventional local projections framework, potentially leading to biased estimates. By combining quantile regression with local projections in a potential outcomes framework, this paper structurally identifies the causal impulse responses of unconditional quantiles of output growth to financial shocks. These impulse responses are state-dependent, but estimating them does not require the problematic inclusion of an endogenous state-variable into the model. Applying our novel framework shows that financial shocks - whether pertaining to credit risk or volatility - cause large output losses, but only in a low growth environments.

Contents

1 Introduction

Relatively small financial shocks that occur when the economy is vulnerable can trigger financial crises causing large and protracted losses to output, consumption and investment. At the same time, similar shocks occurring during "good times" often prove innocuous. Macrofinance literature explains this state-dependence in theoretical models via occasionally binding constraints on financial intermediation (He and Krishnamurthy 2019) and the interaction of shocks with bank balance sheet conditions (Gertler, Kiyotaki, and Prestipino 2019). A large and growing body of empirical literature seeks to estimate state-dependent impulse responses of output growth to financial shocks. However, state-dependence when the state is a function of the outcome variable is particularly problematic in the conventional local projections framework, potentially leading to biased estimates (Gonçalves et al. 2023).

The conventional approach for estimating a state-dependent impulse response uses a variant of local projections that includes a state-dummy interaction. Gonçalves et al. (2023) show that this approach fails when the state-dummy is endogenous. In applied work, it is common to define the state based on the value of the outcome variable at time $t-1$. In particular, this approach has been used to study how the effect of (financial, monetary or government spending) shocks on output depends on whether they occur during recessions or expansions. Unfortunately, defining the state-dummy as a function of the outcome variable means that it is probably endogenous. Moreover, this approach requires the researcher to pick a threshold when constructing the dummy variable. The chosen threshold is often arbitrary and the estimated state-dependent impulse responses are sensitive to this choice. Finally, note that this approach rules out a probable scenario in which the propagation of a time-t shock depends on future states, rather than just on the state at time $t-1$.

In this work we propose a state-dependent estimation approach based on quantile regression. Quantile regression naturally estimates the state-dependent effects of covariates when the state is a mapping from the outcome variable. In a time-series setting, low sample quantiles of the outcome variable map to periods with low realizations of the outcome variable (relative to its sample distribution). Quantile regression estimates quantile specific intercept and slope coefficients, allowing for the effects of covariates to vary along the distribution of the outcome variable. Therefore, the slope coefficients from a quantile regression estimated for a low quantile will capture the effect of the covariates on the outcome variable in these low state periods.

We identify the causal *conditional* and *unconditional* quantile impulse responses (QIRs) by combining generalized quantile regression of Powell (2020) with local projections of Jordà (2005) in a potential outcomes framework. The distinction between effects on conditional versus unconditional quantiles of the outcome variable is crucial for the interpretation of the estimated QIRs as state-dependent impulse responses. Maintaining the desired state-dependent interpretation, while including control variables for the purposes of causal identification, is possible because the generalized quantile regression of Powell (2020) distinguishes between treatment and control variables. In addition, our framework combines other desirable properties. It is based on a definition of QIRs that makes direct comparisons with mean impulse responses possible. Causal identification of QIRs can be achieved using controls (including timing restrictions) and instrumental variables, and does not require a first-stage Structural Vector Autoregression model to identify the structural shocks. Estimation can be done using software packages already available in the Stata SSC archive.

We apply our novel framework to the US data to revisit the effect of financial shocks on economic growth. In both the conditional and unconditional models, we control for; macroeconomic, financial and monetary policy variables to recover the causal effect. Regardless of the conditioning set, low unconditional quantiles of growth always relate to periods of low growth relative to its sample distribution. In contrast, given our set of controls, low conditional quantiles are periods when growth was under-performing given the prevailing macroeconomic, financial and monetary conditions. Our unconditional QIRs can be understood as showing how financial shocks propagate in low versus high growth states. Our findings show that financial shocks - whether pertaining to credit risk or volatility - cause large output losses, but only in low growth environments. Our findings suggest that the heterogeneity in the effects of financial shocks across states of the economy is larger than previously thought. We find persistent output losses of 2% points from a one standard deviation credit risk shock in low growth states, with the median state losses of only 0.5% points, and no losses in the high growth states.

Various methods of identifying and estimating structural QIRs have been proposed in the literature. However, to the best of our knowledge we are the first to develop a framework for the estimation of unconditional QIRs in the presence of control variables. Chavleishvili and Manganelli (2019) achieve identification by imposing timing restrictions on a recursive quantile vector autoregressive model. As such, the QIRs they obtain are defined as responses to quantile shocks (ratio between a demeaned random variable and its quantile). Montes-Rojas (2019) and D. J. Lee, Kim, and Mizen (2021) use the mean-based Vector Autoregression model to identify a structural shock since their multivariate quantile models are reduced-form. Moreover, the QIR proposed by Montes-Rojas (2019) describes the cumulative impact of a series of shocks, not a one-off shock, because persistent realizations of lower (or upper) quantiles are assumed in its construction. Han, Jung, and J. H. Lee (2019) and Jung and J. H. Lee (2022) study QIRs in models where the quantile itself is autoregressive, as in the CAViaR model of Engle and Manganelli (2004). This however is computationally expensive. In the applied literature, Mumtaz and Surico (2015) study the heterogeneity in the transmission mechanism of monetary policy across stages of the business cycle. They estimate the structural QIRs using a quantile autoregressive-distributed lag model of Galvao, Montes-Rojas, and Park (2013) using lags of the monetary policy shock of Romer and Romer as observable structural shocks.

The remainder of the paper is structured as follows. Section 2 introduces the main ideas developed in the paper by way of an illustrative example. In the example, we consider a simplified potential outcomes model for the quantile treatment effect and a simple data generating process with dependence. Section 3 introduces the structural model that allows us to identify the causal *unconditional* and *condtional* QIRs. We focus on causal identification and interpretation of the estimated impulse responses, we close the section by explaining how to compare our QIRs with conventional mean impulse responses. Section 4 contains our empirical findings. We begin by introducing our dataset and explaining the variables capturing credit and volatility risks. We then present and discuss our main findings. Section 5 concludes.

2 Illustrative example

Before introducing our framework for identification of QIRs, we consider a simpler problem of identifying a contemporaneous Quantile Treatment Effect (QTE). Let Y be a scalar outcome variable of interest and let D be a scalar treatment variable. Assume that Y has a potential outcome $Y(d)$ that is the value Y would have taken had treatment status $D = d$ been observed. As such observed $Y \equiv Y(D)$. Further assume that $Y(d)$ has a linear structural quantile function $S(\tau \mid d) = \alpha(\tau) + d\beta(\tau)$, where $\tau \mapsto S(\tau \mid d)$ is non-decreasing on $[0, 1]$ and left-continuous. $S(\tau \mid d)$ describes the quantile of $Y(d)$ obtained by fixing $D = d$ and independently sampling $U(d) \sim Uniform[0, 1]$: $Y(d) = \alpha(U(d)) + d\beta(U(d))$. Our object of interest is:

$$\mathsf{QTE}(\delta) = S(\tau \mid D = d_0 + \delta) - S(\tau \mid D = d_0) = \beta(\tau)\delta \tag{1}$$

which measures how the τ quantile of Y changes when the treatment “dose” is increased by δ from some baseline level d_0. Notice that the assumption of linearity of the structural quantile function means that the QTE is linear in δ and does not depend on the initial level of d_0. Identification of the QTE is challenging in non-experimental settings if there is dependence between the disturbance U and the observed treatment D.

Consider a family of data generating processes that take the form $Y = (\alpha + D\beta)U^*$, where D and U^* are possibly dependent scalar random variables. If $(\alpha + D\beta) > 0 \ \forall D \in \mathbb{D}$, then for any fixed $D = d$, Y is a monotonically increasing transformation of U^* and:

$$S(\tau \mid D = d) = \alpha(\tau) + d\beta(\tau) = [\alpha + d\beta]Q_{U^*}(\tau) \tag{2}$$

where $Q_{U^*}(\tau)$ is the τ quantile of U^*. If U^* is independent of D then a simple quantile regressions of Y on D will recover $\beta(\tau) = \beta Q_{U^*}(\tau)$ and give us an unbiased estimate of the QTE. However, when U^* and D are statistically dependent recovering the QTE becomes difficult, even in cases where the dependence works through observable variables that could be included as additional covariates.

To illustrate the challenges associated with identification of the QTE, consider an example data generating process with dependence on observables. We think of D_t and Y_t as observable stationary time series representing financial conditions and output growth respectively. We think of $E_t \overset{\text{iid}}{\sim} Uniform(0, 0.5)$ and $V_t \overset{\text{iid}}{\sim} Uniform(0, 0.5)$ as unobservable structural shocks to D_t and Y_t respectively.

$$Y_t = \frac{1 + D_{t-1}}{2}[(1 - \rho)V_t + (1 + \rho)Y_{t-1}] \tag{3}$$

$$D_{t-1} = E_{t-1} + V_{t-1} \tag{4}$$

The structural parameter of interest is $\mathsf{QTE}(\delta) = \delta\beta(\tau) = \frac{\delta}{2}Q_{U^*}(\tau)$, where $U_t^* = (1 - \rho)V_t + (1 + \rho)Y_{t-1}$. Note that V_{t-1} is a source of endogeneity of D_{t-1} as it enters into the non-additive error term via Y_{t-1}. We set $\rho = 0.25$. Estimating how the effect of D varies depending on the state U^* is not feasible using Least Squares. To see this we can rewrite equation 3 as:

$$Y_t = \frac{1 - \rho}{2}D_{t-1}V_t + \frac{1 + \rho}{2}Y_{t-1} + \frac{1 + \rho}{2}D_{t-1}Y_{t-1} + \frac{1 - \rho}{2}V_t \tag{5}$$

making it clear that identifying the coefficient on the interaction between D_{t-1} and unobservable V_t is not possible.

This stylized data generating process is meant to capture some prominent features of the likely true data generating process linking financial conditions with output. Namely, financial conditions D affect output Y with a lag, and are themselves a function of the contemporaneous structural financial shock E and structural macroeconomic shock V. Moreover, the effect of lagged financial conditions depends on the present structural macroeconomic shock as well as lagged macroeconomic conditions. Finally, akin to the financial accelerator equation found in numerous macrofinance models following Bernanke, Gertler, and Gilchrist (1999) seminal paper, the relationship is multiplicative.

Simulating this data generating process and attempting to recover the QTE demonstrates the inadequacy of the standard quantile regression for this task. As shown in table 1, quantile regression that ignores the dependence of D fails to recover the QTE. Perhaps more surprisingly adding a lagged value of Y into the regression equation, in an attempt to control for the dependence of D, makes the situation worse. This is because including lagged Y as an additional variable means that the estimated parameter on D only captures the interaction between D and V, rather than the interaction between D and U^*. The generalized quantile regression of Powell (2020) performs better, as it includes the lagged value of Y to account for the dependence of D without changing the interpretation of the estimated parameter. The last two columns are only feasible to estimate if we could observe the structural shock E, in which case using it as an instrument for D further improves identification. Note that we use no knowledge of the functional form of the data generating process (expressed in equations 3 & 4) to recover the QTE using the regressions reported in table 1.

τ	$\mathsf{QTE}(\delta = 1)$	QR(D)	QR(D,L(Y))	GQR	IVQR	IVGQR
0.10	0.84	1.07	1.34	1.07	0.90	0.93
		(0.20)	(0.16)	(0.19)	(0.29)	(0.27)
0.25	1.08	1.32	1.44	1.28	1.12	1.16
		(0.23)	(0.16)	(0.20)	(0.31)	(0.27)
0.50	1.44	1.67	1.48	1.57	1.47	1.53
		(0.31)	(0.16)	(0.24)	(0.39)	(0.29)
0.75	1.93	2.15	1.47	1.99	1.99	2.05
		(0.50)	(0.15)	(0.34)	(0.62)	(0.42)
0.90	2.52	2.71	1.40	2.52	2.66	2.71
		(0.88)	(0.15)	(0.59)	(1.13)	(0.79)

Table 1: Simulation results with $T = 500$ (100 burn-off) and $1,000$ iterations. Estimated coefficients on X_t at five quantile levels $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ reported with Standard Errors in brackets below. Column $\mathsf{QTE}(\delta = 1)$ reports the true value of the targeted causal parameter with $\delta = 1$. Column QR(D) reports results from quantile regression (Koenker and Bassett 1978) of Y_t on X_t. Column QR(D,L(Y)) reports results from quantile regression (Koenker and Bassett 1978) of Y_t on $\{D_t, Y_{t-1}\}$. Column GQR reports results from generalized quantile regression (Powell 2020) of Y_t on D_t as treatment with Y_{t-1} as a control variable. Column IVQR reports results from instrumental variable quantile regression (Chernozhukov and Hansen 2013) of Y_t on D_t using E_t as an instrument. Column IVGQR reports results from generalized quantile regression (Powell 2020) of Y_t on D_t as treatment using E_t as an instrument and Y_{t-1} as a control.

3 Model

Let Y_{t+h} be the outcome variable of interest, measuring the cumulative log growth rate in Y from $t-1$ to $t+h$, where $h \in \{0, 1, 2, \ldots, H\}$ and H is our maximum horizon of interest. Let D_t denote a scalar treatment variable, and collect contemporaneous and lagged control variables into $W_t^\top = (W_{1\,t}, \ldots, W_{p\,t})$.

Define Z_t as an instrumental variable for the treatment variable D_t. We are interested in the dynamic response of Y_{t+h} to D_t over the horizon H, i.e. the impulse response. Specifically we want to estimate the quantile impulse response which is the dynamic response of a given τ quantile of the distribution of Y_{t+h}. Furthermore, we distinguish between the unconditional quantile impulse response (ucQIR) which pertains to the quantile of the unconditional distribution of Y_{t+h}, and the conditional quantile impulse response (cQIR) which pertains to the quantiles of the conditional distribution of $Y_{t+h} \mid W_t$. This is in contrast to a mean impulse response which is the dynamic response of the expectation of the distribution of Y_{t+h}, and for which the distinction between conditional and unconditional expectation is unnecessary due to the Law of Iterated Expectations. In particular, the Law of Iterated Expectations allows us to rewrite a model for the conditional mean $E[Y \mid X] = X\beta$ as $E[E[Y \mid X]] = E[Y] = E[X]\beta$, meaning that $\delta\beta$ captures the effect of increasing X by δ on both the conditional and the unconditional expectation of Y.

To analyze structural quantile impulse responses we build on the potential outcomes framework of Powell (2020) by adapting it to fit our time series setting. In doing so, we follow the notation of Angrist, Jordà, and Kuersteiner (2018). We first present the model in a general form, which allows for identification using instruments, and which distinguishes between control variables in W_t and the treatment variable D_t. Then, we explain how to adapt the framework for identification of the conditional QIR, which makes no distinction between treatment and control variables. We also show how to perform identification by controls (without using the instrumental variable Z_t) in both the unconditional and conditional case. We briefly discuss how to estimate our model and how to obtain the confidence intervals presented in the results section. Lastly, we discuss how to interpret the estimated quantile impulses responses, and how to compare them to mean impulse responses estimated using local projections and Vector Autoregressions.

3.1 Unconditional QIR

The role of the potential outcomes framework presented below, is to make explicit the assumptions that give rise to the moment conditions used to recover the causal effect from observational data.

Assumption 1 (Potential Outcomes). *For a fixed t and h, potential outcome $Y_{t,h}(d)$ is defined as the value that Y_{t+h} would have taken had $D_t = d$ been observed.*

Assumption 2 (Linearity). *$Y_{t,h}(d)$ has a structural quantile function $S_h(\tau \mid d) = \alpha_h(\tau) + d\beta_h(\tau)$, where $\tau \mapsto S_h(\tau \mid d)$ is non-decreasing on $[0, 1]$ and left-continuous.*

Assumption 3 (Conditional Independence). *$Y_{t,h}(d) \perp Z_t \mid W_t$. Potential outcomes $Y_{t,h}(d)$ are conditionally* *(on W_t) independent of the instrument Z_t.*

Assumption 4 (Selection)**.** *$D_t = \omega(Z_t, W_t, V_t)$, for some unknown function ω and an unobservable V_t.*

Assumption 5 (Rank Similarity)**.** *$P[Y_{t,h}(d) \leq S_h(\tau \mid d) \mid Z_t, W_t, V_t] = P[Y_{t,h}(d') \leq S_h(\tau \mid d') \mid Z_t, W_t, V_t]$, $\forall d, d'$.*

Assumption 6 (Observability)**.** *We observe $Y_{t+h} \coloneqq Y_{t,h}(D_t), D_t, W_t, Z_t$.*

Assumption 1 is a standard definition of a potential outcome adapted to the time series setting. Notice that it implicitly assumes that potential outcomes are not allowed to depend on the full treatment path. In other words, each $h-step$ ahead causal effect is only allowed to depend on time t treatment status, excluding dependence on the path of treatment between t and $t+h$. Assumption 2 states that the quantile function that specifies the data generating process behind the potential outcome variable is linear. $S_h(\tau \mid d)$ describes the quantile of the potential outcome variable $Y_{t,h}(d)$ obtained by fixing $D_t = d$ and independently sampling $U_{t,h}(d) \sim Uniform[0,1]$, i.e. $Y_{t,h}(d) = \alpha_h(U_{t,h}(d)) + d\beta_h(U_{t,h}(d))$. The disturbance term $U_{t,h}(d)$ accounts for differences in potential outcomes across observationally equivalent time periods. $U_{t,h}(d)$ can be interpreted as a rank-variable, as by construction events $Y_{t,h}(d) \leq S_h(\tau \mid d)$ and $U_{t,h}(d) \leq \tau$ are equivalent. Assumptions 3, 4 and 5 are key causal identification assumptions that allow us to recover the causal effect of treatment on the quantiles of the outcome variable. Assumptions 3 states that conditional on observables W_t, the instrumental variable Z_t is independent of the potential outcomes. This assumption allows for W_t to provide information about the outcome distribution of $Y_{t,h}(d)$, meaning that the distribution of the unconditional rank variable $U_{t,h}(d)$ can depend on controls W_t. Assumption 4 models the process determining treatment status as a function of the instrument and controls, plus an unobservable term V_t which is a potential source of treatment endogeneity. Assumption 5 states that the distribution of $U_{t,h}(d)$ is the same across treatment states d, this allows for rankings of potential outcomes to vary across treatment states d but only in an asystematic way.

We define the τ-th causal unconditional quantile impulse response to a time-t treatment $d = \delta$ as:

$$\mathsf{ucQIR}_\tau(h, \delta) = S_h(\tau \mid d = d_0 + \delta) - S_h(\tau \mid d = d_0) = \beta_h(\tau)\delta \tag{6}$$

This causal unconditional QIR describes how the quantile of the potential outcome variable $Y_{t,h}(d)$ differs when treatment status is $d = d_0 + \delta$ compared to its value under a counter-factual treatment status $d = d_0$. The fundamental problem of causal inference is that we only observe one realization of $Y_{t,h}(d)$ (i.e.

$Y_{t+h} = Y_{t,h}(D_t)$), and thus we cannot compute the casual treatment effect $Y_{t,h}(d_0+\delta) - Y_{t,h}(d_0)$. This is usually addressed by recovering the average treatment effect $E[Y_{t,h}(d_0+\delta) - Y_{t,h}(d_0)]$. However, this paper seeks to recover the quantile treatment effect, more specifically the ucQIR which is the dynamic quantile treatment effect over horizon H. This is because we are interested in studying the heterogeneity in the effect of treatment across quantiles of the outcome variable, which is "averaged-out" by the average treatment effect.

Before stating the moment conditions used to recover the ucQIR, we reformulate the Theorem 1 from Powell (2020) except for our linear, time series setting.

Theorem 1. *Suppose Assumptions 1-6 hold* $\forall h \in \{0,1,2,\ldots,H\}$. *Then* $\forall h \in \{0,1,2,\ldots,H\}$ *and for each* $\tau \in (0,1)$*:*

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid Z_t, W_t] = \mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid W_t], \tag{7}$$

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau)] = \tau. \tag{8}$$

Equation 7, states that once we condition on controls W_t, the instrument Z_t does not provide additional information about the probability that the outcome is below its quantile function. Equation 8, ensures that the quantile function is correctly scaled. Together, equations 7 & 8 imply that the conditional probability $\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid W_t]$ is allowed to vary based on covariates W_t, but in expectation it is equal to the quantile level τ. Theorem 1 gives us two moment conditions for each $h \in \{0,1,2,\ldots,H\}$:

$$E\{Z_t[\mathbb{I}(Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau)) - P(Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid W_t)]\} = 0 \tag{9}$$

$$E[\mathbb{I}(Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau)) - \tau] = 0 \tag{10}$$

where $\mathbb{I}$ is the indicator function. Estimation using these moment conditions is explained in Powell (2020).

3.2 Conditional QIR

If we are interested in the treatment effect on the conditional quantile of Y we need to estimate a parameter from a different structural quantile function, namely $S_h(\tau \mid d, w) = \alpha_h(\tau) + d\ddot{\beta}_h(\tau) + w^\top\theta_h(\tau)$. $S_h(\tau \mid d, w)$ describes the quantile of the potential outcome variable $Y_{t,h}(d,w)$ obtained by fixing $(D_t = d, W_t = w)$ and independently sampling $\ddot{U}_{t,h}(d,w) \sim Uniform[0,1]$, i.e. $Y_{t,h}(d,w) = \alpha_h(\ddot{U}_{t,h}(d,w)) + d\ddot{\beta}_h(\ddot{U}_{t,h}(d,w)) +$

$w^\top \theta(\ddot{U}_{t,h}(d,w))$. Note that in general $\ddot{\beta}_h(\tau) \neq \beta_h(\tau)$, even if D_t and W_t are independent. This is because $\ddot{\beta}_h(\tau)$ varies along the conditional (on the control covariates in W_t) quantiles of $Y_{t,h}(d,w)$. The mapping between conditional and unconditional quantiles will depend on the conditioning set W_t, which means that the interpretation of $\ddot{\beta}_h(\tau)$ will also depend on what variables are included in W_t. We can think of $U_{t,h}(d) = \omega(\ddot{U}_{t,h}(d,w), W_t)$ for some unknown function ω, where $\ddot{U}_{t,h}(d,w)$ and $U_{t,h}(d)$ refer to the ranks of $Y_{t,h}(d,w)$ and $Y_{t,h}(d)$ respectively.

We define the τ-th causal conditional quantile impulse response to a time-t treatment $d = \delta$ as:

$$\mathsf{cQIR}_\tau(h,\delta) = S_h(\tau \mid d = d_0 + \delta, w) - S_h(\tau \mid d = d_0, w) = \ddot{\beta}_h(\tau)\delta \tag{11}$$

Note that our model is general enough to recover the cQIR. We only need to replace $S_h(\tau \mid d)$ with $S_h(\tau \mid d, w)$ in assumptions 2 & 5, and reformulate assumption 3 as an independence assumption $Y_{t,h}(d,w) \perp Z_t$ that says that potential outcomes $Y_{t,h}(d,w)$ are conditionally independent of Z_t. The implication of this rewritten model is that $\forall h \in \{0,1,2,\ldots,H\}$ and for each $\tau \in (0,1)$:

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\ddot{\beta}_h(\tau) + W_t^\top \theta_h(\tau) \mid D_t, W_t, Z_t] = \mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\ddot{\beta}_h(\tau) + W_t^\top \theta_h(\tau) \mid Z_t] = \tau. \tag{12}$$

3.3 Identification by controls

So far we have focused on causal identification by instrumental variable Z_t. However, the model can be easily rewritten for identification by controls, by simply setting $Z_t = D_t$. In that case, assumption 4 is trivially satisfied. The observability assumption 6 loses Z_t as it no longer plays any role in the model, we only need to observe $Y_{t+h} \coloneqq Y_{t,h}(D_t), D_t, W_t$.

In the unconditional model identified by controls, assumption 3 becomes $Y_{t,h}(d) \perp D_t \mid W_t$, which means that endogeneity of D_t is addressed by assuming that treatment status is conditionally (on controls) independent of potential outcomes. The implication of this rewritten model for the ucQIR is that $\forall h \in \{0,1,2,\ldots,H\}$ and for each $\tau \in (0,1)$:

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid D_t, W_t] = \mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau) \mid W_t], \tag{13}$$

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t\beta_h(\tau)] = \tau. \tag{14}$$

In the conditional model identified by controls, assumption 3 becomes $Y_{t,h}(d) \perp D_t, W_t$, i.e. we assume treatment status and controls are independent of potential outcomes. The implication of this rewritten model for the cQIR is that $\forall h \in \{0, 1, 2, \ldots, H\}$ and for each $\tau \in (0, 1)$:

$$\mathbb{P}[Y_{t+h} \leq \alpha_h(\tau) + D_t \ddot{\beta}_h(\tau) + W_t^\top \theta_h(\tau) \mid D_t, W_t] = \tau. \tag{15}$$

3.4 Estimation and confidence intervals

Our framework naturally lends itself to estimation of the QIRs by local projections. For a fixed quantile level τ, QIRs are defined as a set of $H + 1$ coefficients, which can be estimated as the set of coefficients on the treatment variable D_t from local projections on Y_{t+h}. In particular, depending on whether we are estimating the ucQIRs or the cQIRs and whether we use the instrument Z_t for identification or not, we need to save the coefficients on D_t from local projections on Y_{t+h} recovered by estimating (separately for each $h \in \{0, 1, 2, \ldots, H\}$ and for each quantile τ) :

	Identification by Controls	Identification by Instrumental Variables
cQIR	QR of Koenker and Bassett (1978)	IV-QR of Chernozhukov and Hansen (2013)
ucQIR	GQR of Powell (2020)	GQR of Powell (2020)

We calculate confidence intervals using block bootstrap. This procedure entails re-sampling the data by randomly drawing blocks of m consecutive observations, using blocks which start at indexes $1, \ldots, T - m + 1$, before re-estimating the model B times using these pseudo-samples. The confidence intervals are then based on the distribution of the estimated parameters across the B repetitions of the procedure. Block bootstrap is more appropriate than stationary bootstrap for time series applications, because re-sampling blocks of observations preserves the temporal dependence in the data.

3.5 Connection with mean impulse responses

One strength of our chosen definition of the quantile impulse response is its similarity with the definition of a mean impulse response that many researchers are already familiar with. The mean impulse response of Y_t to an impulse is often defined as:

$$IR(h, \delta) = \mathbb{E}[Y_{t+h} \mid D_t = d_0 + \delta] - \mathbb{E}[Y_{t+h} \mid D_t = d_0]. \tag{16}$$

By comparing the above with the definitions of the ucQIR and the cQIR expressed in equations 6 & 11, the similarity should be self-evident. Compared to alternative definitions of the quantile impulse response used in the literature, our definition has the following desirable properties making comparisons with mean impulse responses more direct; it captures the impact of a one-off shock rather than a series of shocks, it does not rely on quantile specific shocks, it does not require a first-stage Structural Vector Autoregression model to identify a structural shock.

Having said that, a word of caution is in order when dealing with cumulative quantile impulse responses. To calculate cumulative impact on growth in the level of the variable of interest (e.g. Industrial Production IP_t) using local projections, the outcome variable is usually transformed to $Y_{t+h} = log(IP_{t+h}) - log(IP_{t-1})$. This is also the transformation used in this paper. This transformation is innocuous in the case of the mean impulse response as linearity of the expectations operators implies:

$$IR(h,\delta) = \sum_{s=0}^{h} \{\mathbb{E}[log(IP_{t+s}) - log(IP_{t+s-1}) \mid D_t = d_0 + \delta] - \mathbb{E}[log(IP_{t+s}) - log(IP_{t+s-1}) \mid D_t = d_0]\} \quad (17)$$

meaning that the effect on average cumulative growth is equal to the sum of the effects on the consecutive between period average growth rates. Importantly, the $Y_{t+h} = log(IP_{t+h}) - log(IP_{t-1})$ transformation is not as innocuous in the case of quantile impulse responses, as generally $Q_{A+B}(\tau) \neq Q_A(\tau) + Q_B(\tau)$ unless the random variables A and B are comonotonic. For example, the effect on the median annual growth rate will not generally equal to the sum of the effects on the 12 consecutive median monthly growth rates. This has implications for how we should interpret cumulative quantile impulse responses. In particular, we should read the quantile impulse response at horizon h as describing how the τ quantile of the h periods ahead distribution of cumulative growth is affected by a time-t shock of size δ to D_t.

Plagborg-Møller and Wolf (2021) show that under appropriate assumptions the local projection and Vector Autoregression impulse responses are equal, up to a constant of proportionality. The presence of the constant of proportionality comes from the fact that the implicit local projection innovation (after controlling for the other right-hand side variables) does not have unit variance, unlike the innovations in a Vector Autoregression model. Plagborg-Møller and Wolf (2021) provide an expression for this constant of proportionality, which makes it possible to compare the magnitude of the impulse responses estimated using local projection and Vector Autoregression frameworks. Similarly to the local projection mean impulse response, the QIR estimated by our model should be interpreted as a response to a δ change in the treatment variable D_t, rather than a response to a unit innovation to the treatment variable D_t. Therefore, if we want to compare the QIRs with mean impulse responses from a local projections, we can simply ignore the constant of proportionality, and if we want to compare them with impulse responses from a Vector Autoregression we can set δ equal to the constant of proportionality.

4 Empirical Results

4.1 Data

Our monthly dataset covers the US economy from February 1986 to August 2021 (T=427). All of the data we use in the paper is publicly available, with majority of it contained in the FRED-MD database published by the St. Luis Fed. We use monthly data to benefit from a larger sample size. Due to the unavailability of monthly GDP we focus on Industrial Production (IP) as the dependent variable. This is a natural choice, as IP accounts for the bulk of the variation in output over the course of the business cycle.

Throughout, the dependent variable Y_{t+h} will be defined as the h-months cumulative log growth rate $Y_{t+h} = 100 * [log(IP_{t+h}) - log(IP_{t-1})]$. We multiply the log growth rates by 100 to interpret the QIR in terms of percentage points. We Z-score normalize the treatment variable D_t to interpret the QIRs as responses to a one standard deviation increase in treatment "dose".

The first treatment variable D_t we consider measures exogenous movements in credit risk. We will refer to this variable as credit risk and we define it as the first difference of the monthly Excess Bond Premium (EBP) of Gilchrist and Zakrajšek (2012), i.e. $D_t = EBP_t - EBP_{t-1}$. The EBP is the residual credit spread that cannot be explained by the usual counter-cyclical movements in expected defaults, movements which account for less than one-half of the variation in corporate bond credit spreads (an empirical observation known as the "credit spread puzzle"). Credit Spread is the difference in yield of a corporate bond versus a treasury bond promising the same cash-flow, and as such it measures the credit risk premium demanded by investors.

The second treatment variable D_t we consider measures volatility surprises in the equity markets. We will refer to it as volatility risk and we define it as the difference between realized and implied volatility of the S&P500 index. The volatility implied in an option's price is widely regarded as the option market's forecast of future return volatility over the remaining life of the relevant option. If option markets are efficient, implied volatility should be an efficient forecast of future volatility, it should subsume the information contained in all other variables in the market information set in explaining future volatility. If implied volatility is an unbiased forecast of realized volatility then a regression of the form $realized_t = \alpha + \beta implied_t + \epsilon_t$ should yield $\alpha = 0$, $\beta = 1$. Assuming that the efficient market hypothesis holds ($\alpha = 0$, $\beta = 1$), $D_t = realized_t - implied_t$ should give us a time series of volatility that was unexpected by the financial markets (Christensen and Prabhala

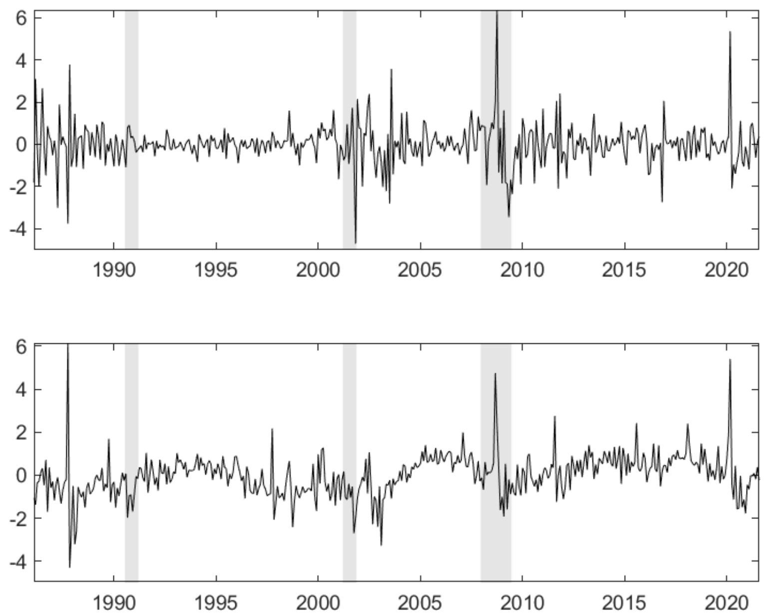

Figure 1: Time series of the two treatment variables D_t. Top panel plots the normalized first difference of the monthly Excess Bond Premium. Bottom panel plots the normalized difference between the realized and implied volatility of the S&P500 index. Grey bands indicate NBER recession dates.

1998).

In the literature, it is often assumed that real variables are slow moving while financial and monetary variables adjust quickly. For instance, Gilchrist and Zakrajšek (2012) order the EBP after macroeconomic variables but before financial markets and monetary policy variables in a Structural Vectror Autoregression model used to study the effects of EBP shocks on the macroeconomy. We follow the same logic by ordering our variables as follows: {consumption growth, investment growth, industrial production growth, inflation, financial variable Dt, S&P500 monthly return, change in the 10 year the ten-year (nominal) Treasury yield, change in the effective (nominal) federal funds rate}. This ordering implies that W_t must include the contemporaneous values of the four variables ordered before the treatment variable D_t. Additionally, to control for the broad state of the economy in the recent past, we include the first two lags of all eight variables contained in our ordering in W_t. In short, our timing restriction assumption allows that the treatment variable D_t adjusts within the period to consumption growth, investment growth, industrial production growth and inflation, but adjusts with a one month's lag to the stock market return, changes of the Treasury yields and changes to the Fed's funds rate.

4.2 Results

Throughout, we focus on three quantiles $\tau \in \{0.1, 0.5, 0.9\}$ representing low growth, median growth and high growth states respectively. These refer to quantiles of h-periods ahead cumulative growth. To understand the corresponding growth rates in our sample, figure 2 plots the unconditional quantiles of cumulative IP growth over the horizon scaled by the number of months that the growth is cumulated over.

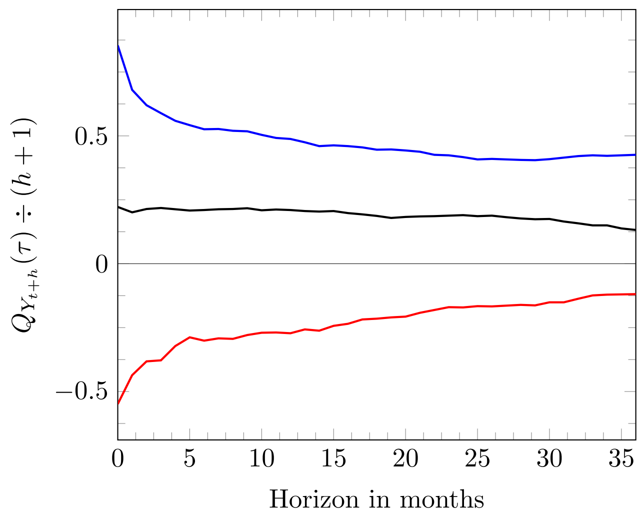

Figure 2: Cumulative Industrial Production growth quantiles divided by the number of months growth is cumulated over, i.e. $Q_{Y_{t+h}}(\tau) \div (h+1)$. Plotted for $\tau = \{0.1, 0.5, 0.9\}$. Monthly Industrial Production growth quantiles are: $Q_{Y_t}(0.1) = -0.557$, $Q_{Y_t}(0.5) = 0.220$, and $Q_{Y_t}(0.9) = 0.853$.

Figure 3 shows the recovered ucQIRs of industrial production to a one standard deviation increase in credit risk. The upper-left panel in figure 3 plots the ucQIRs for the three quantiles on the same axis. It is clear that the response in the low growth state is much more pronounced than in the other two states. This is a feature of the data and not of the model, as nothing is restricting the responses of lower quantiles to be lower than those of the upper quantiles. These findings suggest economically large and statistically significant (at 90% confidence level) growth losses of about 2% points when a credit risk shock propagates in a low growth environment. The losses in the median state are considerably smaller at around 0.5% points. The response in high growth states is muted and not persistent.

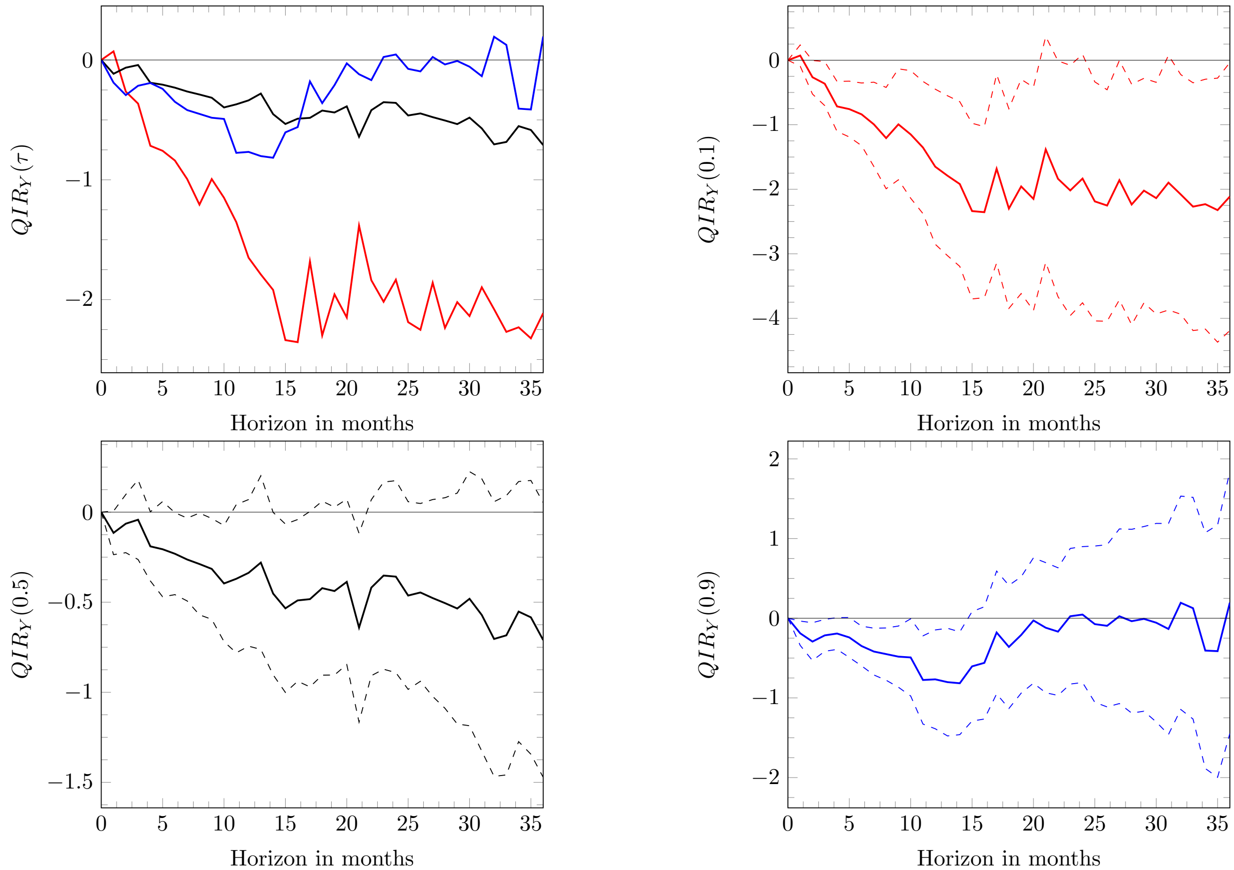

Figure 3: Unconditional QIR with 90% Block Bootstrap CIs (block size $m = 7$, bootstrap repetitions $B = 500$). $\tau = 0.1$ in red, $\tau = 0.5$ in black, $\tau = 0.9$ in blue. $\delta = 1$. Effect size should be read as cumulative loss to Industrial Production growth (in % pts.) from 1 std. dev. increase in the credit risk shock.

Comparing figure 4 to figure 3 suggests that the relationship between volatility risk and growth is like the relationship between credit risk and growth. The timing, magnitude and asymmetry in the quantile effects are similar following increases in volatility risk and credit risk. This is true despite the fact that the sample correlation coefficient between the two treatment variables is very low ($\hat{\rho} = 0.096$).

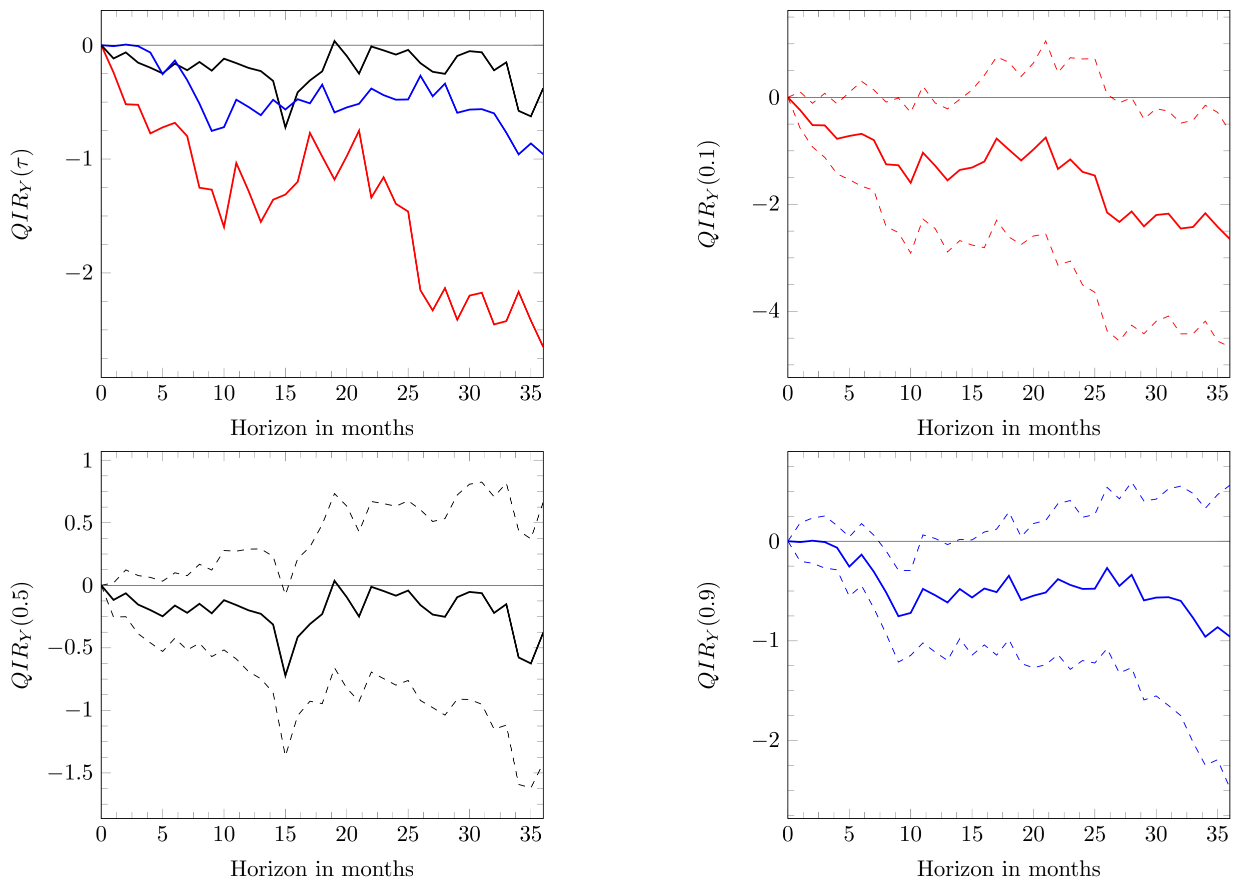

Figure 4: Unconditional QIR with 90% Block Bootstrap CIs (block size $m = 7$, bootstrap repetitions $B = 500$). $\tau = 0.1$ in red, $\tau = 0.5$ in black, $\tau = 0.9$ in blue. $\delta = 1$. Effect size should be read as cumulative loss to Industrial Production growth (in % pts.) from 1 std. dev. increase in the volatility risk shock.

Lastly, we compare the findings presented above with findings obtained using the conventional state-dummy local projections approach. In particular, we estimate:

$$Y_{t+h} = H_{t-1}[\alpha_h(1) + \beta_h(1)D_t + \theta_h(1)W_t^\top] + (1 - H_{t-1})[\alpha_h(0) + \beta_h(0)D_t + \theta_h(0)W_t^\top] + \varepsilon_{t+h}, \quad (18)$$

where H_{t-1} is a dummy variable taking value 1 if $Y_{t-1} > 0$ (i.e. if monthly IP growth was positive in pre-treatment period) and 1 otherwise. Estimating the above by Least Squares gives us two state-dependent impulse responses $\hat{\beta}_h(1)$ and $\hat{\beta}_h(0)$. We also estimate the model without the dummy interaction:

$$Y_{t+h} = \alpha_h + \beta_h D_t + \theta_h W_t^\top + \varepsilon_{t+h}, \quad (19)$$

which gives us the mean impulse response estimate $\hat{\beta}_h$. Figure 5 compares the results obtained using this framework with the ucQIRs. Recall that the states in the case of ucQIR refer to the quantiles of h-periods ahead cumulative growth, and as such are distinct from the definition of the state used in the state-dummy local projections approach. Therefore, the differences in estimated responses among these two alternative frameworks are to be expected. Having said that, it is possible to make a few observations.

Firstly, it is reassuring to see that the ucQIRs for the median are broadly similar to the estimated mean impulse responses. In addition, for the credit risk specification our mean and median impulse responses are of similar magnitude and shape to the impulse responses of real GDP to EBP shocks estimated by Gilchrist and Zakrajšek (2012) using a Structural Vector Autoregression model with quarterly data. Secondly, shocks that propagate in low-states (captured by the ucQIR for $\tau = 0.9$, red line on left panels) are approximately 50% more harmful than shocks that occur following a period of negative industrial production growth (captured by $\hat{\beta}_h(0)$, red line on right panels). Thirdly, our quantile regression based methodology estimates more asymmetric responses across states, with responses in low growth states being clear outliers compared to high and median growth state responses.

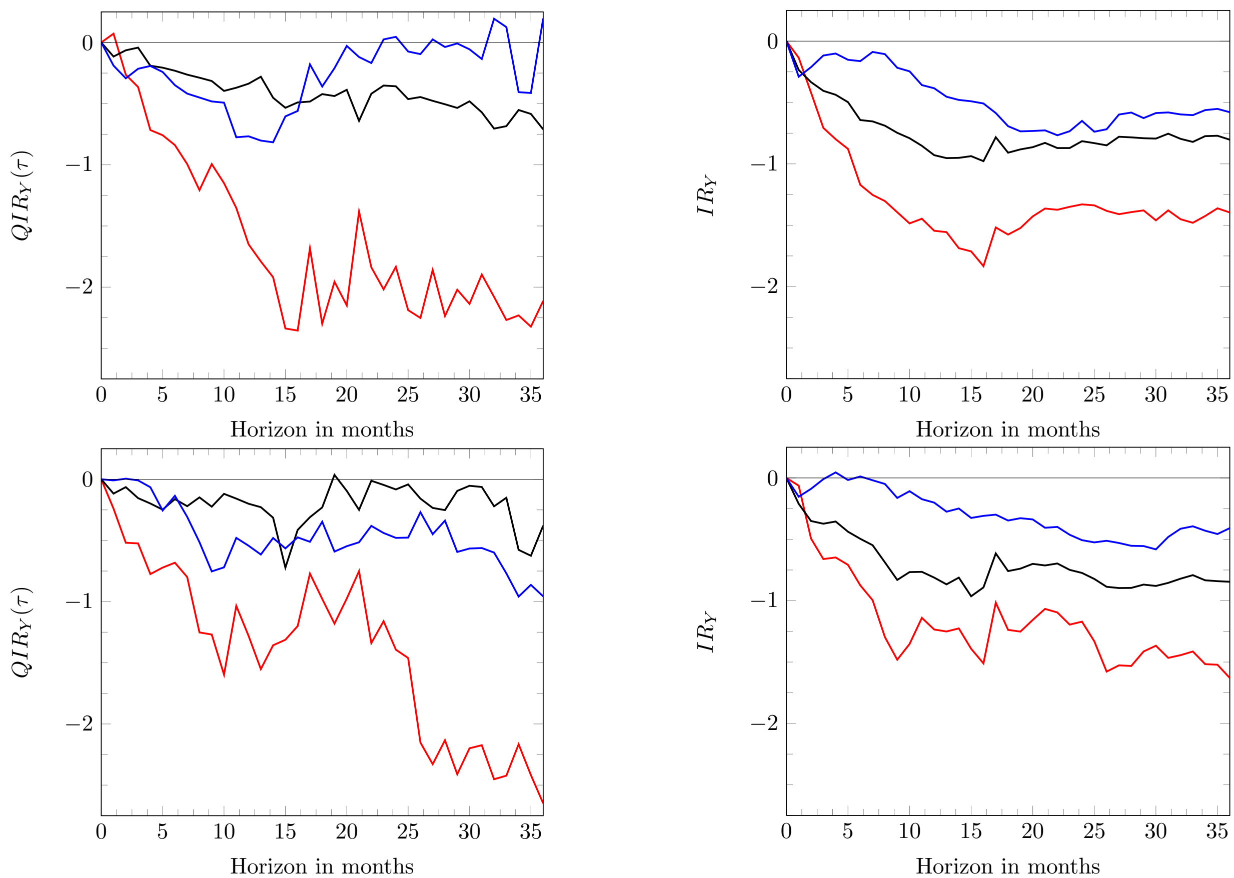

Figure 5: Panels in the top row show responses to credit risk, bottom row panels show responses to volatility risk. Left panels present ucQIRs for $\tau = 0.1$ in red, $\tau = 0.5$ in black, $\tau = 0.9$ in blue. Right panels present state-dependent mean impulse responses. State = 1 if first lag of monthly IP growth was positive when shock hit, 0 otherwise. Blue line shows the IR conditional on state = 1 ($\hat{\beta}_h(1)$), red line shows the IR conditional on state = 0 ($\hat{\beta}_h(0)$). Black line, mean impulse response estimated ignoring state-dependence ($\hat{\beta}_h$).

5 Conclusion

The relationship between financial shocks and the macroeconomy is complex and volatile. Understanding the theoretical mechanisms driving this relationship is the core ambition of the macrofinance literature. Theory tells us that the state of the economy when adverse financial shocks occur determine how the shocks will propagate. This idea has spurred many empirical papers which test it in the data. We offer new evidence based on a novel causal framework for state-dependence. Our empirical findings corroborate prior work but also offer new insights. We think that the key distinction between our methodology and the conventional approach is that we allow future states to influence the shape of the state-dependent impulse response. This allows us to capture a higher degree of state-dependence. Although our model is tailored to the empirical application of interest, it could be readily applied for studying other important macroeconomic questions. For instance, measuring the state-dependent size of the fiscal multiplier and the transmission of monetary policy across stages of the business cycle.

6 Bibliography

Angrist, Joshua D., Òscar Jordà, and Guido M. Kuersteiner (2018). "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited". In: *Journal of Business & Economic Statistics* 36.3, pp. 371–387.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist (1999). "Chapter 21 The financial accelerator in a quantitative business cycle framework". In: vol. 1. Handbook of Macroeconomics. Elsevier, pp. 1341–1393.

Chavleishvili, Sulkhan and Simone Manganelli (2019). "Forecasting and stress testing with quantile vector autoregression". In: *ECB Working Paper* (2330).

Chernozhukov, V. and C. Hansen (2013). "Quantile Models with Endogeneity". In: *Annual Review of Economics* 5.1, pp. 57–81.

Christensen, B.J. and N.R. Prabhala (1998). "The relation between implied and realized volatility." In: *Journal of Financial Economics* 50.2, pp. 125–150.

Engle, Robert F. and Simone Manganelli (2004). "CAViaR: Conditional autoregressive value at risk by regression quantiles". In: *Journal of Business and Economic Statistics* 22 (4).

Galvao, Antonio F., Gabriel Montes-Rojas, and Sung Y. Park (Apr. 2013). "Quantile Autoregressive Distributed Lag Model with an Application to House Price Returns". In: *Oxford Bulletin of Economics and Statistics* 75 (2), pp. 307–321.

Gertler, Mark, Nobuhiro Kiyotaki, and Andrea Prestipino (May 2019). "A Macroeconomic Model with Financial Panics". In: *The Review of Economic Studies* 87.1, pp. 240–288.

Gilchrist, Simon and Egon Zakrajšek (June 2012). "Credit Spreads and Business Cycle Fluctuations". In: *American Economic Review* 102.4, pp. 1692–1720.

Gonçalves, Silvia et al. (2023). "State-dependent local projections". In: *FRB of Dallas Working Paper*.

Han, Heejoon, Whayoung Jung, and Ji Hyung Lee (2019). "Estimation and Inference of Quantile Impulse Response Functions by Local Projections: With Applications to VaR Dynamics". In: *SSRN Electronic Journal*.

He, Zhiguo and Arvind Krishnamurthy (Oct. 2019). "A Macroeconomic Framework for Quantifying Systemic Risk". In: *American Economic Journal: Macroeconomics* 11.4, pp. 1–37.

Jordà, Òscar (2005). "Estimation and inference of impulse responses by local projections". In: *American Economic Review* 95 (1).

Jung, Whayoung and Ji Hyung Lee (2022). "Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance". In: *Advances in Econometrics*.

Koenker, Roger and Gilbert Bassett (Jan. 1978). "Regression Quantiles". In: *Econometrica* 46 (1).

Lee, Dong Jin, Tae-Hwan Kim, and Paul Mizen (2021). "Impulse response analysis in conditional quantile models with an application to monetary policy". In: *Journal of Economic Dynamics and Control* 127, p. 104102.

Montes-Rojas, Gabriel (2019). "Multivariate Quantile Impulse Response Functions". In: *Journal of Time Series Analysis* 40 (5), pp. 739–752.

Mumtaz, Haroon and Paolo Surico (2015). "The transmission mechanism in good and bad times". In: *International Economic Review* 56.4, pp. 1237–1260.

Plagborg-Møller, Mikkel and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impulse Responses". In: *Econometrica* 89 (2).

Powell, David (Dec. 2020). "Quantile Treatment Effects in the Presence of Covariates". In: *The Review of Economics and Statistics* 102.5, pp. 994–1005.

7 Appendix

7.1 Conditional QIR

Here we present the results for the *conditional* QIRs identified using controls. We control on the same set of variables W_t as in our main results. The difference is that here we are not distinguishing between treatment and control variables. We use equation 15 to get $H + 1$ moment conditions:

$$E\{\mathbb{I}[Y_{t+h} \leq \alpha_h(\tau) + D_t\ddot{\beta}_h(\tau) + W_t^\top \theta_h(\tau)] - \tau\} = 0. \tag{20}$$

We estimate this using quantile local projections on Y_{t+h} with a standard quantile regression of Koenker and Bassett (1978). As such impulse responses for low quantiles τ now refer to conditional on W_t quantiles of Y_{t+h}. Put simply, low quantiles now refer to conditionally on macro/financial/monetary conditions low growth states. Note that depending on the prevailing conditions, these "low" states may occasionally map to periods of unconditionally high growth. Therefore, the interpretation of QIRs as state-dependent is not as straight forward as in the unconditional presented in the main text.

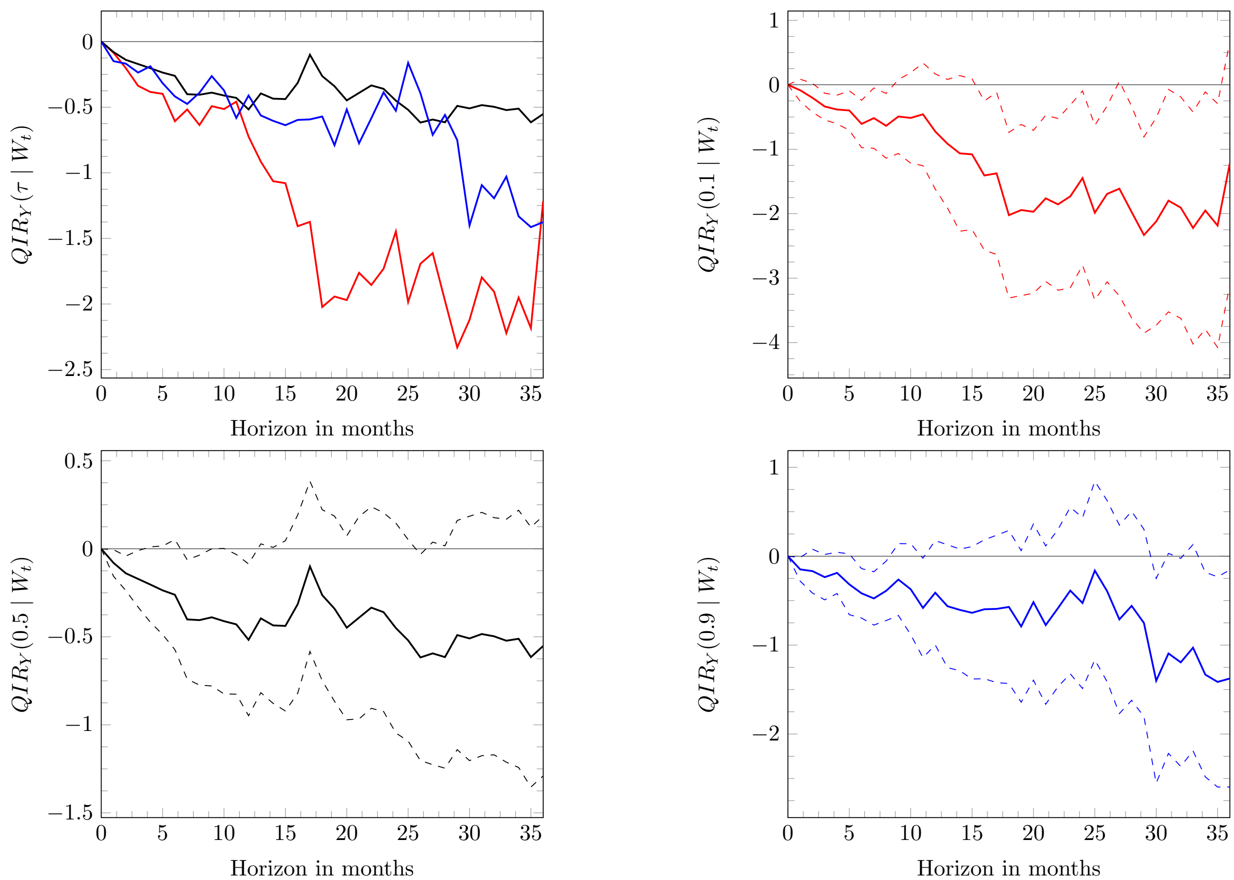

Figure 6: Conditional QIR with 90% Block Bootstrap CIs (block size $m = 7$, bootstrap repetitions $B = 1,000$). $\tau = 0.1$ in red, $\tau = 0.5$ in black, $\tau = 0.9$ in blue. $\delta = 1$. Effect size should be read as cumulative loss to Industrial Production growth (in % pts.) from 1 std. dev. increase in the credit risk.

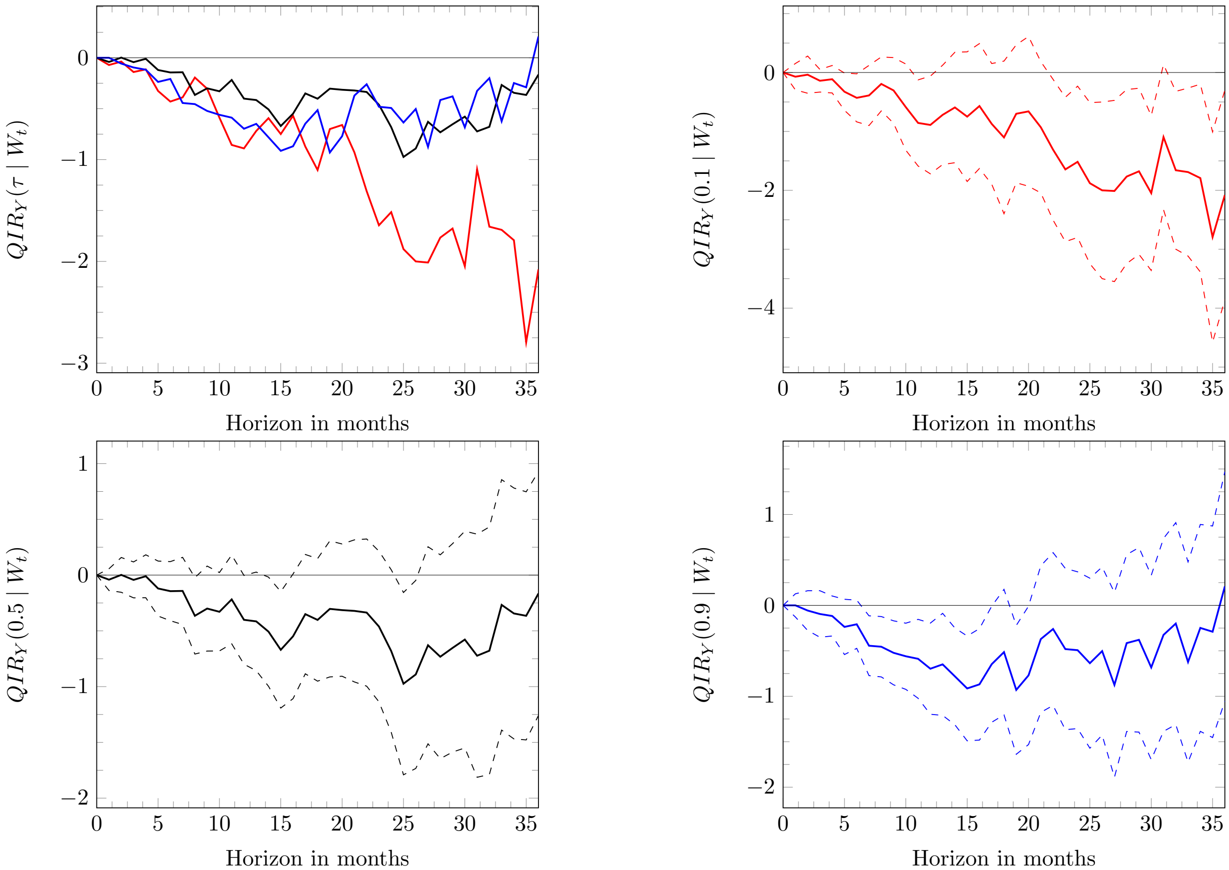

Figure 7: Conditional QIR with 90% Block Bootstrap CIs (block size $m = 7$, bootstrap repetitions $B = 1,000$). $\tau = 0.1$ in red, $\tau = 0.5$ in black, $\tau = 0.9$ in blue. $\delta = 1$. Effect size should be read as cumulative loss to Industrial Production growth (in % pts.) from 1 std. dev. increase in the volatility risk.