A Nonlinear Dynamic Factor Model for Financial and Macroeconomic Data*

Pablo A. Guerrón Quintana† Alexey Khazanov‡ Molin Zhong§

January 12, 2026

Abstract

We propose a nonlinear dynamic factor model to examine the impact of external shocks and internal forces on macroeconomic and financial data fluctuations. This model permits nonlinear dynamics, enables asymmetric, state-dependent, and size-dependent responses to shocks, and generates time-varying volatility and asymmetric tail risk behavior. We find evidence of nonlinear dynamics in a U.S. application on measuring the credit cycle. The nonlinear factor stimulates credit growth during booms but hinders recovery after crises, with shocks having longer, amplified effects during credit crunches. An extended model that separates first- and second-order effects of the factor reveals that credit cycles exhibit state dependence in which shocks that spur growth during modest credit conditions trigger sharp busts during excessive expansions, thereby predicting how credit conditions can flip from boom to crisis.

Keywords— Asymmetric dynamics, credit cycle, second-order factor, state dependence, tail risk

*We thank Todd Clark, Taeyoung Doh, Thorsten Drautzburg, Domenico Giannone, Matteo Iacopini, Dongho Song, and two anonymous referees for insightful comments and suggestions. We also benefited from comments by seminar participants at several venues. All errors are ours. This paper does not necessarily reflect the views of the Federal Reserve System or its Board of Governors. We acknowledge the use of ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google) as drafting and editing aids.

†Boston College, email: pguerron@gmail.com.

‡The Hebrew University of Jerusalem, email: alexey.khazanov@mail.huji.ac.il.

§Board of Governors, email: Molin.Zhong@frb.gov.

1 Introduction

What are the drivers of financial and macroeconomic variables – external shocks or endogenous propagation? This timeless question becomes particularly salient during downturns, when the scope for economic policy is most consequential. Dynamic factor models (DFMs) have played a central role in addressing this question by summarizing the comovement of large macroeconomic datasets (Stock and Watson, 2016). Since their introduction, DFMs have relied largely on vector autoregressive (VAR) representations that impose linear relationships in both the factor dynamics and the measurement equation. However, substantial evidence documents pervasive nonlinearities in macroeconomic variables (Baker et al., 2016, Fernández-Villaverde et al., 2015b, Justiniano and Primiceri, 2008). Financial data, in particular, exhibit abrupt changes during crises (Gilchrist and Zakrajsek, 2012, Ludvigson et al., 2021). In this paper, we study macroeconomic and financial data using a DFM that incorporates such nonlinearities.

The nonlinearities introduced in our DFM capture several features emphasized in the recent literature that existing factor models largely overlook. The resulting nonlinear dynamic factor (NLDF) model enables us to examine nonlinear state dynamics during periods of elevated volatility such as the Global Financial Crisis (GFC) and to construct point forecasts and predictive densities in the presence of nonlinearities. Importantly, we study these mechanisms within a unified framework.

Our specification of the nonlinear factor model is inspired by the pruned second-order state-space (2nd SS) representation used to approximate nonlinear dynamic stochastic general equilibrium (DSGE) models (Kim et al., 2008, Andreasen et al., 2017). As emphasized by Fernandez-Villaverde et al. (2016), the pruned solution captures key nonlinear features while preserving stationarity, a property we exploit here. We reinterpret the 2nd SS framework in the context of a DFM, allowing the factor to evolve according to a nonlinear state equation, and explicitly analyze how the model's nonlinear structure delivers identification of the factor's dynamics. These features render the Kalman filter inappropriate, motivating our use of particle filtering methods. We propose two Bayesian estimation approaches: one based on particle Gibbs with ancestor sampling (Lindsten et al., 2014) and another using Metropolis-Hastings combined with a bootstrap particle filter.

Nonlinear state dynamics generate rich implications for impulse response functions (IRFs) and predictive densities. Although nonlinearities are introduced at the factor level, they are transmitted to observables through the measurement equation. The resulting IRFs exhibit three notable properties: asymmetry (same-sized positive and negative shocks produce different responses), state dependence with respect to initial conditions, and size dependence across shock magnitudes. Exogenous shocks that have small effects in normal times may become amplified during crises. The model also has implications for higher-order moments. Despite homoskedastic and normally distributed innovations, nonlinear dynamics induce non-normal unconditional state distributions. Predictive densities display time-varying volatility and asymmetric tail risks, consistent with evidence documented by Adrian et al. (2019).

Because these nonlinear dynamics operate through latent factors, careful attention to identification is essential. We show that, under a linear measurement equation, the introduction of nonlinear (second-order) dynamics in the latent state does not fundamentally alter the classic identification challenges of linear DFMs: the sign and scale of the factors remain unidentified without normalization. We formally establish this result and demonstrate that standard identification strategies, such as fixing factor loadings for selected observables, continue to apply in the nonlinear setting. Importantly, because the second-order component is a deterministic function of past first-order factors with the same factor loadings in the benchmark model, identifying the first-order component is sufficient to identify the nonlinear component as well. We then study an extended specification in which the first- and second-order components load separately onto observables and show that this additional flexibility introduces new identification challenges, requiring additional normalizations to separately pin down the nonlinear component. Monte Carlo experiments confirm that our proposed identification schemes successfully recover the true parameters and latent factors, and that ignoring nonlinear dynamics leads to systematic distortions in estimated persistence and factor behavior.

We apply the NLDF model to study the roles of external shocks and internal propagation in the U.S. credit cycle. We estimate a nonlinear credit cycle as the common component of U.S. credit growth across nonfinancial business, household, financial, and public sectors. This analysis is motivated by extensive work emphasizing the role of credit and leverage in nonlinear shock amplifi-

cation (Bernanke et al., 1999, Brunnermeier and Sannikov, 2014, Schularick and Taylor, 2012). The nonlinear factor captures the slow buildup and sharp contractions common across credit sectors, highlighting the importance of a slow-moving second-order component that boosts credit growth from the mid-1990s until the GFC and whose collapse contributes to the weak post-crisis recovery. The effects of shocks are state dependent: shocks during credit crunches are more persistent and amplified than those during booms. Negative shocks raise predictive uncertainty, generating larger downside tail risks than upside risks.

We further estimate an extension of the NLDF model that allows first- and second-order components to have different weights in the measurement equation. Doing so clearly separates the contribution of the first- and second-order components to credit growth dynamics. We find that the second-order component becomes the dominant cyclical factor capturing common credit movements while exhibiting much stronger nonlinearities than in the baseline model. The extended model reveals important state dependence where the second-order component's response to shocks changes sign depending on whether the first-order component is positive or negative at the time of impact. This mechanism helps explain credit boom-bust dynamics, where shocks that stimulate credit growth when credit conditions are modest trigger sharp declines with increased volatility when credit expansions become excessive.

Related Literature. The literature has explored several nonlinear extensions of factor models, including Markov switching, time-varying parameters, and stochastic volatility. To the best of our knowledge, however, no existing work allows for second-order dynamics in the state equation.

Our work is closest to two recent contributions. Aruoba et al. (2017) propose a quadratic autoregressive (QAR) model that incorporates quadratic terms and GARCH features using a pruned representation. Their focus is univariate and directly on observables, whereas our framework is multivariate and factor-based. Gorodnichenko and Ng (2017) exploit second-order DSGE solutions to extract factors associated with levels and volatility. In contrast, our model has a functional formulation that is analogous to the pruned nonlinear DSGE solution (Andreasen et al., 2017), uses likelihood-based estimation rather than principal components, and is designed to study higher-order moments, state-dependent IRFs, and predictive densities.

More broadly, our work contributes to several strands of the extensive factor model literature reviewed by Stock and Watson (2016). Related work includes regime-switching factor models (Chauvet, 1998), DFMs with time-varying parameters and stochastic volatility (Del Negro and Otrok, 2008), and Markov-switching factor models applied to economic activity and credit sentiment (Leiva-León et al., 2022). Additional advances include methods for handling missing data (Banbura and Modugno, 2014), nonparametric diffusion models (Shintani, 2005), semiparametric factor structures (Chen et al., 2021), and DFMs with structural breaks (Cheng et al., 2016). Our work also relates to regime-switching approaches to the ELB (Aruoba et al., 2021), studies allowing for non-Gaussian shocks (Gourieroux et al., 2019), and the growing literature on tail risks and distributional asymmetries (Adrian et al., 2019).

The rest of the paper is organized as follows. The next section discusses the NLDF model. We motivate the factor model by connecting it to the pruned solution of a nonlinear DSGE model. In Section 3, we highlight the novel implications of the nonlinear DFM for moments, IRFs, and predictive densities. Section 4 discusses our empirical application on the credit cycle. Some concluding remarks are in the final section.

2 The Nonlinear Dynamic Factor Model

In this section, we introduce the NLDF model and then discuss the motivation for specifying our factor dynamics. Next, we discuss factor identification under a linear measurement equation specification. Finally, we close the section by presenting our estimation algorithms.

2.1 Model Specification

We consider the following NLDF model:

$$\text{Measurement: } y_t = \mathcal{G}(f_t) + \eta\epsilon_t \tag{1}$$

$$\text{Factor dynamics: } f_t = \mathcal{H}(f_{t-1}) + \sigma\nu_t. \tag{2}$$

Here, ϵ_t is an $N \times 1$ vector of $iid \sim N(0, I_N)$ innovations, ν_t is a $K \times 1$ vector of $iid \sim N(0, I_K)$ innovations, y_t is an $N \times 1$ vector of observed variables, and f_t is the $K \times 1$ underlying factor. $\mathcal{G}(\cdot)$ and $\mathcal{H}(\cdot)$ are functions discussed below. In addition, we assume that the $\mathcal{H}(\cdot)$ function is at least twice differentiable. η is an $N \times N$ diagonal matrix of standard deviations, and σ is a $K \times K$ matrix that is the square root of a variance-covariance matrix. The additive measurement error assumption is for ease of exposition. Our framework can easily handle multiplicative errors, like in Hwang (1986), or nonadditive errors.

For empirical applications involving data such as unemployment, gross domestic product (GDP) growth, or inflation rates, a typical desirable feature of the factor process is stationarity. A generic $\mathcal{H}(\cdot)$ function, estimated using a limited data range, may imply explosive dynamics of the factor and, thus, of the observables. To avoid this problem, we use the pruned state equation (Kim et al., 2008, Andreasen et al., 2017) that has easily verifiable stationarity conditions.

To make the analysis more concrete, consider an NLDF model in which the measurement equation is linear in a single underlying factor and we take a pruned second-order approximation to the function $\mathcal{H}(\cdot)$. We take the single factor specification as a leading case, although we also explore extensions to multiple factors. Let f_t denote the underlying factor and f_t^f and f_t^s its first- and second-order terms such that $f_t = c + f_t^f + f_t^s$. Then the pruned system is

$$
\begin{aligned}
y_t &= G f_t + \eta \epsilon_t \\
&\begin{cases}
f_t = c + f_t^f + f_t^s \text{ – first- and second-order factors} \\
f_t^f = h_x f_{t-1}^f + \sigma \nu_t \\
f_t^s = h_x f_{t-1}^s + \frac{1}{2} h_{xx} \left(f_{t-1}^f \right)^2 .
\end{cases}
\end{aligned}
\tag{3}
$$

The first-order term follows the same process as a linear DFM with the persistence parameter governed by h_x. The exogenous shocks ν_t perturb the first-order term on impact. The second-order term depends on the square of the lagged first-order term, with h_{xx} modulating the importance of this relationship. The second-order term also has persistence determined by h_x. The exact structure of this process is the second-order solution to a DSGE model that prevents explosive paths (Andreasen et al., 2017).

This structure has two main attractive properties. First, the model allows for rich nonlinearities due to the presence of the second-order term. Specifically, the model can generate asymmetric, state-dependent, and size-dependent IRFs. Moreover, the model can generate time-varying volatility through the state dependence. We will illustrate these properties in detail in Section 3. Second, the model has easily verifiable stationarity conditions. As long as $|h_x| < 1$, the model is stationary.

The quadratic term h_{xx} generates a non-zero mean for the factor, even if the shocks have zero mean. This property can be verified by applying the expectation operator on the second-order term in Equation 3. We include a constant c in the factor's law of motion to adjust the overall factor to have zero mean in our applications, although this parameter may alternatively be estimated.

2.2 Motivating the Nonlinear Dynamic Factor Model

Our nonlinear factor model brings together two strands of literature: one on solving nonlinear DSGE models and one showing the importance of quadratic factors in macro and financial applications.

Connection to Theoretical Macroeconomic Models. Our time series model is closely connected to the nonlinear solution of a DSGE model, which we view as an important strength of our framework. Nonlinear DSGE models have rapidly grown in popularity, spurred on by an ample body of empirical research that documents that the U.S. economy has nonlinear features such as stochastic volatility (Justiniano and Primiceri, 2008, Fernández-Villaverde and Rubio-Ramírez, 2007, Bloom, 2009, Fernández-Villaverde et al., 2015a), time-varying monetary policy (Fernández-Villaverde et al., 2015b), and the zero lower bound on short-term interest rates (Fernandez-Villaverde et al., 2015, Gust et al., 2017b, Wu and Xia, 2016).

Our NLDF model is the direct time series analogue of the pruned second-order perturbation DSGE solution. Suppose the DSGE model has only one state variable (denoted f_t). Then, its

dynamic equation is approximated by

$$\begin{aligned} \text{State equation: } f_t &= f_t^f + f_t^s \\ f_t^f &= h_1 f_{t-1}^f + \sigma \nu_t \\ f_t^s &= h_1 f_{t-1}^s + \frac{1}{2} h_2 \left(f_{t-1}^f \right)^2. \end{aligned} \tag{4}$$

Here, h_1 and h_2 are coefficients and ν_t is a normally distributed innovation. Comparing Equations 3 and 4, we see that their structures are the same. The key difference is that in the DSGE model solution, h_1 and h_2 are known given the deep parameters of the model. The h_2 parameter summarizes the nonlinearities that arise in policy functions when we solve the model to second order. These nonlinearities include asymmetric, state-dependent, and size-dependent responses to shocks (Andreasen et al., 2017). In our time series model, by contrast, the corresponding parameters are estimated from the data.

Importance of Quadratic Factors. The second-order dynamics specified in our model can also be motivated by an econometric literature investigating the importance of quadratic factors in predictive regressions. There are two chief approaches. The first specifies squared factors in the regression (Ludvigson and Ng, 2007, Bai and Ng, 2008):

$$y_{t+h} = \beta_0 + \beta_1 F_t + \beta_2 F_t^2 + \epsilon_{t+h}, \tag{5}$$

where y_{t+h} is the series of interest at horizon h and F_t is a principal component extracted in an initial step from a large cross section of possible predictors. A second one, termed squared principal components, extracts factors from a cross section that contains both the level and the square of many series (Bai and Ng, 2008).

These approaches have been applied in several contexts, including in modeling the risk-return relationship in stock markets (Ludvigson and Ng, 2007), predicting inflation (Bai and Ng, 2008), and modeling the conditional mean of many macro and financial series (Jurado et al., 2015). In these applications, the quadratic components consistently play an important role, illustrating the importance of allowing for this form of nonlinearity.

Our factor model also contains second-order terms, but we specify this nonlinearity in the factor dynamics. Through the measurement equation, the second-order terms, which are discounted sums of the squares of first-order factors, enter as determinants of the observable variables. In contrast, the literature on quadratic factors is less concerned about the dynamics of the factors and is instead primarily focused on their value in predicting observable variables, with the notable exception of Gorodnichenko and Ng (2017). Importantly, in comparison to other important classes of nonlinear DFMs, such as time-varying parameter DFMs (Del Negro and Otrok, 2008) or Markov-switching DFMs (Chauvet, 1998), we believe our modeling strategy hews more closely to this literature.

2.3 Extensions

While our benchmark model follows the structure laid out in Andreasen et al. (2017), we also consider an extension of our baseline NLDF model where the first- and second-order components have different weights in the measurement equation. The extended model is as follows:

$$\begin{aligned} y_t &= G_1 f_t^f + G_2(c + f_t^s) + \eta\epsilon_t \\ &\begin{cases} f_t^f = h_{x,1} f_{t-1}^f + \sigma\nu_t \\ f_t^s = h_{x,2} f_{t-1}^s + \frac{1}{2} h_{xx} \left(f_{t-1}^f\right)^2. \end{cases} \end{aligned} \tag{6}$$

We introduce a new factor loading G_2 that determines how the observable loads onto the second-order component. As the effects of the first- and second-order factors become less tightly linked, we allow for two different levels of persistence $h_{x,1}$ and $h_{x,2}$. Stationarity in this model requires that $|h_{x,1}| < 1$ and $|h_{x,2}| < 1$.

In Appendix B, we also discuss potential ways in which the model could be extended to study data that demand a richer factor structure such as higher-order representations of the state equation and a multidimensional state. In addition, we explore a variation of the benchmark model which allows a nonlinear relation between the observables and the dynamic factor (see Appendix Section I).

2.4 Identification of Factors

Given the nonlinearities modeled in the latent factor, a natural benchmark case is for the measurement equation to be linear, as in Equation 3. The linear measurement equation allows us to prove that the latent factors are not identified and an additional normalization is needed.

Benchmark Model. Consider the multivariate version of the NLDF model with K factors, which we show in Equation 7:

$$
\begin{aligned}
&y_t = Gf_t + \eta\epsilon_t \\
&\begin{cases}
f_t = c + f_t^f + f_t^s \text{ – first- and second-order components} \\
f_t^f = h_x f_{t-1}^f + \sigma\nu_t \\
f_t^s = h_x f_{t-1}^s + \frac{1}{2} h_{xx} vech\left(f_{t-1}^f \otimes f_{t-1}^f\right)
\end{cases}
\end{aligned} \tag{7}
$$

where $vech$ is the half-vectorization operator and h_{xx} is a $K \times \tilde{K}$ matrix with $\tilde{K}$ the number of unique elements in the Kronecker product. This representation allows for an interaction of the first-order components. In the case of 2 nonlinear factors: $f_{1,t-1}^f \times f_{2,t-1}^f$.

Proposition 2.1. The sign and scale of the factors in the nonlinear factor model in Equation 7 are not identified.

The proof is in the Appendix Section A. In our application that uses a single factor, we fix $G_1 = 1$, which corresponds to the named factor approach in the DFM literature (Stock and Watson (2016)). In general, the usual identification conditions for linear dynamic factor models, such as those presented in Bai and Wang (2015), continue to apply. We can continue to use this identification strategy because the first- and second-order components enter additively into the measurement equation. Therefore, the presence of the second-order components does not affect the identifiability of the first-order components. Since the second-order components are deterministic functions of past first-order components and they load onto the measurement equation with the same factor loadings, they inherit identification once first-order terms are identified.

Separate Weights on the First- and Second-Order Components. Consider the multivariate version of the extended NLDF model with separate weights on the first- and second-order components with K factors:

$$\begin{aligned} y_t &= G_1 f_t^f + G_2 \left(c + f_t^s\right) + \eta \epsilon_t \\ &\begin{cases} f_t^f = h_{x,1} f_{t-1}^f + \sigma \nu_t \\ f_t^s = h_{x,2} f_{t-1}^s + \frac{1}{2} h_{xx} vech\left(f_{t-1}^f \otimes f_{t-1}^f\right). \end{cases} \end{aligned} \tag{8}$$

Proposition 2.2. The sign and scale of the first- and second-order components of the factors in the nonlinear factor model in Equation 8 are not identified.

The proof is in the Appendix Section A. To identify the first-order components, we can use the strategies in Bai and Wang (2015). To identify the second-order components, we can impose that a $K \times K$ block of G_2 is the identity matrix. The difference relative to the previous case is that identifying the first-order components is insufficient to identify the second-order components. This is because the second-order components load separately onto the measurement equation. Therefore, even when the first-order components are identified, there still is a full rank matrix that can affect the identification of the second-order components.

2.5 Estimation Algorithm

We use Bayesian methods to estimate the model. Given the model's nonlinearity, we rely on particle filtering methods to approximate the likelihood (Särkkä, 2013). Specifically, our empirical illustration uses a Gibbs sampler combined with a particle smoother, which delivers smoothed estimates — the most accurate estimates of the factor given all available data. By exploiting ancestor sampling (Lindsten et al., 2014), this algorithm achieves good mixing properties even with relatively few particles, on the order of hundreds. This computational efficiency has become increasingly important as particle filtering methods have gained popularity in economics, aided by recent advances such as differentiable particle filters (Brady et al., 2025) and accessible toolboxes for coding Graphics Processing Units (Deng et al., 2023). The method has proven successful in estimating complex nonlinear state space models with many state variables Fernández-Villaverde

and Rubio-Ramírez (2007), Bocola (2016), Gust et al. (2017a), and Herbst and Schorfheide (2019). Further details about our estimation algorithm and its computational implementation can be found in Appendix Section E.

Monte Carlo. In Appendix Section F, we conduct some Monte Carlo exercises to study estimation performance. First, we show that if the data-generating process is the nonlinear factor model itself, our estimation strategy can recover the true parameter values. Second, we find that the likelihood implied by a linear factor model is below the likelihood from the nonlinear model if the data were generated from our NLDF model. By ignoring the nonlinear dynamics, the linear model tends to estimate an excessively persistent linear factor. Finally, we verify our identification strategies in a two factor model and an extension of our model with separate factor loadings on the first- and second-order components. We view these identification exercises as an additional contribution to the existing literature.

3 Properties of the Nonlinear Dynamic Factor Model

We now move on to some key properties generated by the benchmark NLDF model. Our focus is on the latent factor with the understanding that these properties propagate through to the observables via the measurement equation. We focus on three novel features that our NLDF model brings to the table: asymmetric responses to positive versus negative shocks, state-dependent responses, and size-dependent responses. These properties of IRFs were previously discussed in the context of solution methods to DSGE models in Andreasen et al. (2017), but we find it instructive to review them here. A linear DFM cannot deliver these types of IRFs.

3.1 A Useful State-Space Form

We can write the factor dynamics in a useful state-space form, first presented in Andreasen et al. (2017), to make analytical progress on the model's implications for the moments of the factors. To facilitate the exposition, let us continue to assume a one-dimensional factor. Then one can write the factor dynamics as follows:

$$f_t = c + f_t^f + f_t^s \tag{9}$$

$$\begin{pmatrix} f_t^f \\ f_t^s \\ \left(f_t^f\right)^2 \end{pmatrix} = \begin{pmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}h_{xx} \\ 0 & 0 & h_x^2 \end{pmatrix} \begin{pmatrix} f_{t-1}^f \\ f_{t-1}^s \\ \left(f_{t-1}^f\right)^2 \end{pmatrix} + \begin{pmatrix} \sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma^2 & 2\sigma h_x \end{pmatrix} \begin{pmatrix} \nu_t \\ \nu_t^2 \\ f_{t-1}^f \nu_t \end{pmatrix}. \tag{10}$$

For ease of notation, we write the state space as follows:

$$z_t = Az_{t-1} + B\zeta_t, \tag{11}$$

where $z_t = [f_t^f, f_t^s, \left(f_t^f\right)^2]'$ and $\zeta_t = [\nu_t, \nu_t^2, f_{t-1}^f \nu_t]'$. The matrices A and B contain the corresponding parameters in the state equation. It is possible to show that the innovations ζ_t are intertemporally uncorrelated – in other words, $E\left[\zeta_{t+i}\zeta_{t+j}\right] = 0$ for $i \neq j$. In this section, when we discuss conditional moments at time t, the conditioning set is the past history of factors $\{f_t^f, f_t^s, f_{t-1}^f, f_{t-1}^s, ...\}$.

3.2 First Moment Dynamics

We begin by discussing first moment dynamics. We emphasize four parts: the persistence of the factor, asymmetry in the impulse response to a positive versus negative shock, state dependence in the response, and size dependence in the response.

The general formula for conditional mean dynamics at horizon h is $E_t z_{t+h} = A^h z_t + \sum_{i=0}^{h-1} A^i [0, 0, \sigma^2]'$. There are two points to mention from this equation. First, the h_x term determines the persistence property of the entire system, as is expected. If $|h_x| < 1$, the conditional mean responses converge. Second, the ν_t^2 innovation produces a nonzero long-run mean for $\left(f_t^f\right)^2$.

The overall factor has persistence from both the first- and second-order factors. The first-order factor, f_t^f, has a persistence equal to h_x. The second-order factor, f_t^s, has a persistence greater than h_x. The second-order factor has more persistence because while f_t^s follows an autoregressive process with parameter h_x, its "innovation" depends on $\left(f_{t-1}^f\right)^2$, which itself is persistent with parameter h_x^2. As long as $h_x > 0$, which is usually the case for macroeconomic and financial data, the second-order factor is more persistent than the first-order one.

The ν_t^2 component of the innovation generates an asymmetric response to a positive versus negative shock because while a positive shock increases ν_t – and therefore f_t^f – and a negative shock decreases it, both a positive and negative shock increase ν_t^2. The important parameter governing the direction of the asymmetry is h_{xx}. This parameter governs how $\left(f_{t-1}^f\right)^2$ relates to f_t^s and therefore how the effects of ν_t^2 pass through to the overall factor. If the sign of h_{xx} is positive, then a positive shock increases f_t^f and the response of ν_t^2 increases f_t^s. A negative shock decreases f_t^f, but the response of ν_t^2 still increases f_t^s. The effect is reversed if the sign of h_{xx} is negative.

The impulse response is state dependent, which comes from the $f_{t-1}^f \nu_t$ term in the innovation. The sign of f_{t-1}^f determines the effect of a shock to ν_t on $\left(f_t^f\right)^2$. Additionally, the magnitude of f_{t-1}^f determines the amount of time-varying volatility, which we discuss more in the next subsection.

The ν_t^2 term also creates size dependencies in the response to a shock, meaning that a two standard deviation shock does not generate double the responses of a one standard deviation shock. This effect follows straightforwardly from the quadratic transformation. This feature of our model is important because it can generate strong amplification to shocks during downturn episodes.

3.3 Volatility Dynamics

The model generates time-varying volatility via the state dependence inherent in the second-order factor. The magnitude of f_{t-1}^f modulates the effect of ν_t on $\left(f_t^f\right)^2$. A larger value of f_{t-1}^f means that the same-sized shock generates a larger response in $\left(f_t^f\right)^2$. The quadratic component of the model responds by more the further f_{t-1}^f is away from 0.

$$V_t\left(z_{t+h}\right) = \sum_{i=0}^{h-1} A^i B V_t\left(\zeta_{t+h-i}\right) B' \left(A'\right)^i \tag{12}$$

$$V_t\left(\zeta_{t+h}\right) = \begin{pmatrix} 1 & 0 & h_x^{h-1} f_t^f \\ 0 & 2 & 0 \\ h_x^{h-1} f_t^f & 0 & E_t\left(\left(f_{t+h-1}^f\right)^2\right) \end{pmatrix} \tag{13}$$

The formula for the h-step-ahead conditional variance of the system is shown in Equations 12

and 13. The latter equation shows that the $f_{t-1}^f \nu_t$ term generates time-varying volatility in the system. The conditional variance from this term depends on $\left(f_{t-1}^f\right)^2$ and therefore has a persistence of h_x^2. As Equation 12 shows, the conditional variance of z_t at various horizons is then a discounted sum of the conditional variance of ζ_t from $t+1$ up through $t+h$.

3.4 Relationship Between First and Second Moments

Our model generates a non-zero correlation between conditional first and second moments. This relationship can be seen by noticing that $\left(f_{t-1}^f\right)^2$ simultaneously determines the conditional mean of f_t^s and the conditional volatility of shocks to $\left(f_t^f\right)^2$. Simultaneous movements in mean and volatility are an important mechanism identified in the growth-at-risk literature to generate asymmetric tail risk behavior (Adrian et al., 2019).

More formally, we can examine the conditional covariances between f_t^f and $\left(f_t^f\right)^2$ and f_t^s. These expressions can be derived from the general formulas in Equations 12 and 13. The formula for the covariance between the first-order factor and its square at horizon h is:

$$Cov_t\left(f_{t+h}^f, \left(f_{t+h}^f\right)^2\right) = 2\sigma^2 h_x^h \frac{1-h_x^{2h}}{1-h_x^2} f_t^f. \tag{14}$$

As suggested by the earlier discussion on conditional mean dynamics, the sign of f_t^f is important. Specifically, if f_t^f is positive and $h_x > 0$, the conditional covariance between the two terms is positive as well and vice versa. To understand why, consider the one step ahead case. Note that the shock ν_{t+1} determines f_{t+1}^f, while ν_{t+1}^2 and $f_t^f \nu_{t+1}$ determine $\left(f_{t+1}^f\right)^2$, conditional on knowing f_t^f. The shocks ν_{t+1} and ν_{t+1}^2 are uncorrelated, so any nonzero covariance must come from $f_t^f \nu_{t+1}$. When f_t^f is positive, then further increases in the first-order factor increase $\left(f_{t+1}^f\right)^2$ and the correlation between the first-order factor and its square is positive, whereas when f_t^f is negative, further decreases in the first-order factor increase $\left(f_{t+1}^f\right)^2$ and the correlation is negative. The same intuition holds true for longer horizons as well due to the positive autocorrelation of f_t^f. Therefore, if f_t^f is positive, f_{t+i}^f is expected to remain positive and therefore continue producing a positive comovement between the first-order factor and its square. The overall covariance takes into account all of the shocks that occur between $t+1$ and $t+h$.

The conditional covariance between the second-order factor and the first-order factor squared is 0 at one step ahead, as the second-order factor at time $t+1$ is predetermined given time t information. The conditional covariance becomes nonzero at two steps ahead with a value at horizon h given by

$$Cov_t\left(f^s_{t+h}, \left(f^f_{t+h}\right)^2\right) = h_x Cov_t\left(f^s_{t+h-1}, \left(f^f_{t+h}\right)^2\right) + \frac{1}{2}h_{xx}Cov_t\left(\left(f^f_{t+h-1}\right)^2, \left(f^f_{t+h}\right)^2\right). \quad (15)$$

The two terms are tightly related because the second-order factor directly loads onto past values of the first-order factor squared. As can be seen by examining Equation 15, the conditional covariance is determined by two components. The first part of the expression accounts for the persistence in f^s_{t+h} as its value at time $t+h$ depends on its value at $t+h-1$. The $t+h-1$ term also covaries with $\left(f^f_{t+h}\right)^2$ because f^s_{t+h-1} depends on $\left(f^f_{t+h-2}\right)^2$ and further lags of the squared first-order factor. The second part of the expression depends on h_{xx} and the conditional one-step-ahead autocovariance of $\left(f^f_{t+h-1}\right)^2$. If $h_{xx} > 0$, then increases in the squared first-order factor generate increases in the second-order factor, and the conditional covariance is positive. The effects are reversed if $h_{xx} < 0$. The autocovariance term appears because the time $t+h$ value of the second-order factor loads onto the time $t+h-1$ value of the squared first-order factor, so intertemporal dynamics play a role.

In Appendix Section C, we discuss the unconditional correlations between the first and second moments. There is an unconditional dependence between the second-order factor and volatility which induces a relationship between level and volatility. In Appendix Section D, we use a calibrated version of the model to illustrate the IRF and distributional responses to shocks.

4 The Nonlinear Credit Cycle

Since the GFC, economists have once again taken a close look at the importance of credit growth for macroeconomic fluctuations (Schularick and Taylor, 2012). Excessive credit buildups often precede financial crises, and leverage can further amplify shocks. Moreover, it is not enough to focus on one credit sector, but instead a broad monitoring framework is needed (Adrian et al.,

2015). For instance, Mian et al. (2017) emphasize the importance of household debt to GDP as a predictor of lower GDP growth and higher unemployment worldwide. Corporate leverage may lead to distorted investment decisions due to debt overhang effects (Gomes et al., 2016). Financial-sector leverage can amplify shocks via the financial accelerator and binding borrowing constraints (Bernanke et al., 1999, Gertler and Karadi, 2011). Finally, as discussed in Jorda et al. (2016), high levels of public debt tend to prolong the pain of private-sector deleveraging. Taking center stage in these studies is the importance of credit growth.

Our empirical application investigates the importance of a common component in real credit growth in the United States across the nonfinancial business, household, financial, and public sectors from 1952:Q1 through 2021:Q4. Economic theory suggests the potential importance of nonlinearities in determining the dynamics of credit growth (Brunnermeier and Sannikov, 2014, Guerrieri and Iacoviello, 2017). We view our model as one avenue to check how important the nonlinear dynamics are in the data without needing to resort to a fully specified structural model.

For our benchmark results, we estimate a one-factor version of our nonlinear factor model (Equation 3). We use a particle Gibbs sampling algorithm with 100 particles and take 1.5 million draws from the posterior distribution, burning in the first $600,000$. To form our posterior distribution, we take every 300th draw for a total of $3,000$ draws. For posterior distributions of IRFs and distributional moments, which require heavier computation, we use $1,000$ draws of the parameters. Appendix Sections E and H contain further details about the estimation such as the prior specification and data.

4.1 Historical Credit Cycle Estimates

Our estimates provide evidence of a nonlinear factor that we call the credit cycle. For identification purposes, we fix the factor loading for nonfinancial business credit growth at 1. The factor positively loads onto the household and financial sectors, with 100% of draws above 0 in both cases. Indeed, the posterior median of the factor loading on household credit is 1.3, with nearly all draws above 1, while the posterior median of the loading on financial credit is around 1. The factor, therefore, is heavily informed by the common cyclical co-movement of the three private credit growth series. The factor, however, also plays a role in understanding the public credit growth dynamics.

It has a factor loading of -0.2 on the government credit growth series, with nearly all draws less than 0. Therefore, the factor broadly captures the correlation dynamics we document in the data in Appendix Section H.

[Figure 1 about here]

Figure 1 shows the smoothed factor estimates. In the top panel, the dashed line is the posterior median of the nonlinear factor estimates along with the 68% credible bands. The credit cycle factor was strong throughout the 1960s before the recession in 1969. It then rebounded before collapsing again during the mid-1970s recession, with similar dynamics repeating again in the late 1970s to early 1980s. The frequency of the credit cycle lengthened afterward, with a robust expansion in the 1980s before declining again in the late 1980s and early 1990s with the savings and loan crises. Following that episode was a prolonged expansion through the 1990s and 2000s before the collapse in the GFC. The recovery from the GFC was especially slow, with the credit factor still below its mean even more than 10 years after the recession. This finding points to a secular stagnation in financial markets, that is, the GFC resulted in a significant and permanent change in financial markets. Indeed, we can see that the stagnation captured by the factor arises from the dynamics of credit in the household and financial sectors. Interestingly, our estimated nonlinear credit cycle looks quite similar to the credit market sentiment index reported by Leiva-León et al. (2022), who rely on a Markov-switching approach.

The solid line and shaded areas are the corresponding movements of the first-order factor only for comparison purposes. These are the counterfactual estimates of the factor if we had set h_{xx} to 0 across all of the draws, holding all else equal. The bottom panel shows the estimates of the second-order factor adjusted to have 0 mean. The nonlinear component of the model was significantly positive starting in the 1970s, providing a boost to credit growth. It then declined to negative territory in the late 1980s during the savings and loan crisis. The factor again turned positive for a 15-year stretch beginning in the early 1990s until the GFC, when the second-order factor swung heavily negative. This negative swing contributed to the sluggish recovery of credit growth post-crisis.

4.2 Importance of the Nonlinear Factor

How important is the nonlinear factor when modeling the credit cycle? We answer this question in two ways. First, we investigate the state-dependent effects of shocks conditioning on three periods: the credit boom in the mid-2000s, the bust in the late 2000s and early 2010s, and a mixed case in the late 1980s. We find the last period listed particularly interesting, as it had a positive first-order and overall credit factor but a negative second-order factor. This finding is in contrast to the first two periods, in which the first- and second-order factors had the same signs. Second, we look at the standard deviation and tail risk effects of shocks. It is important to reemphasize that in a linear DFM, shocks do not have state-dependent nor higher-order moment effects.

The details about how we generate the IRFs are in Appendix Section D. In Appendix Section H, we also look at the unconditional distribution of the nonlinear model compared with a counterfactual one with only the first-order factor active.

State-Dependent and Asymmetric Effects of Shocks. Two key aspects of the nonlinear credit cycle are state-dependent and asymmetric responses to shocks. Figure 2 shows the responses to one standard deviation positive and negative shocks to the credit cycle factor. The dashed line and shaded areas are the responses from the nonlinear model, while the solid line and shaded areas are the responses from the linear model. The first row shows the effects of a positive shock, whereas the second row shows the effects of a negative shock. The columns condition on the smoothed state estimates of three different periods: a boom period in the mid-2000s, the bust after the GFC in 2010, and a mixed case leading into the early 1990s recession.

[Figure 2 about here]

During the credit boom period, where both the first- and second-order factors were positive, the expected path of the credit factor actually behaves similarly to the linear-only model. The persistence of the first-order factor, governed by the h_x parameter, has a posterior mean of 0.92, with the nonlinear factor showing similar intertemporal dynamics. As we move to the credit crunch period in 2010, the first- and second-order factors both are negative, which generates a response to the shock that is more persistent and with a slight hump shape in the initial quarters. There is

a change in the conditional volatility of the shock when compared with the credit boom period as well, with the magnitudes of the responses to the same-sized shock larger in the quarters after its realization. These findings are consistent with the unconditional distribution results, which show a negative relationship between the level of the credit factor and its conditional volatility. The final column of the figure shows a mixed period before the early 1990s recession. The smoothed first-order factor was positive but declining, while the second-order factor became negative. This combination of states leads to a response to the credit factor shock that dies out more quickly compared with both the credit boom and crunch states. This response is true both for a positive and negative shock.

In summary, the nonlinear model exhibits evidence of state dependence in the responses to a shock. The results are in line with theoretical predictions as well. When the credit cycle is strong, a credit factor shock behaves approximately linearly. These times correspond to periods of slack borrowing constraints and easy credit (Guerrieri and Iacoviello, 2017). Times immediately after credit crunches generate amplification and persistence as borrowing constraints tighten. Our empirical results suggest that dynamics in the data are consistent with these theories.

We also comment briefly on the asymmetry in the responses to positive versus negative shocks. Across all of the periods, evidence indicates that a negative shock generates a larger and more persistent response when compared with a positive shock. A negative shock leads to a response approximately 10% larger in magnitude when compared with a positive one. In Section H of the Appendix, we present results on the difference in the magnitudes between positive and negative shocks. We also search for evidence of size dependencies in the response to a shock and in unreported results find little evidence of this mechanism at play for either a positive or negative shock.

Higher-Order Moment Effects of Shocks. Figure 3 shows the higher-order moment effects of shocks. For these results, we condition on the credit crunch state, although many of the qualitative features we discuss apply to the other times as well.

[Figure 3 about here]

A positive shock leads to an increase in the mean and a decline in the volatility of the credit factor predictive distribution, as seen on the first two columns of the figure. As we move to the last

column, we see the effects that these shocks have on the tail risk of the predictive distribution. The shortfall increases more than the longrise does because the increase in mean and decrease in volatility both lead to the lower tail of the distribution shifting leftward. In contrast, these effects partially cancel each other out on the upper end of the distribution, generating the more muted response.

The bottom row shows the response to a negative shock. The responses flip in sign, with the shock generating an increase in the volatility. Both the shortfall and longrise decline, with the decline in the shortfall still greater than the decline in the longrise. Taken together, these results suggest that a credit cycle shock produces larger moves in downside risk relative to upside risk. An adverse credit cycle shock lowers the factor on average, and it also increases the risk of particularly large declines due to an increase in volatility. In contrast, a positive shock increases the factor, on average, and further decreases the risk of large declines due to a decline in volatility.

4.3 Comparison to Gorodnichenko and Ng (2017)

As discussed in the literature review, Gorodnichenko and Ng (2017) outline an alternative multi-step procedure to extract level and volatility factors. We employ their methodology to extract factors from our credit growth data. Figure 4 compares the first- and second-order components of our one-factor nonlinear model with those of the Gorodnichenko-Ng approach. The upper panel shows that our first-order component correlates strongly with the level factor. In contrast, our second-order component differs significantly from the volatility factor. This is visually clear during the late 1980s and the years following the 2008 financial crisis. We believe that the differences are due to three reasons. First, our approaches are motivated by different solution methods, Gorodnichenko and Ng (2017) builds off of the Benigno et al. (2013) approximation while our work uses the pruned perturbation solution of Andreasen et al. (2017). Second, Gorodnichenko and Ng (2017) use principal components methods and a multi-step estimation procedure while we estimate the nonlinear system jointly using likelihood-based methods. Finally, Gorodnichenko and Ng (2017) allows for independent variation in the second moment factor, while our methodology ties the second-order factor to squares of the past first-moment factor.

[Figure 4 about here]

In Appendix Section H, we present additional results from this analysis, including measures of goodness of fit. Overall, we conclude that our NLDF model is complementary to Gorodnichenko and Ng (2017)'s framework.

4.4 Extension to Separate Weights for First- and Second-Order Components

We estimate a one-factor model that allows for separate weights for the first- and second-order components in the measurement equation, as shown in Equation 6. To identify the model, we set $G_{1,1} = 1$ and $G_{1,2} = 1$ for nonfinancial business credit growth. Our estimation specifications are the same as for the benchmark model and the prior distributions can be found in Appendix Section H. We impose that the persistence of the first- and second-order components are the same.

There are three main takeaways from the estimation. First, the second-order component becomes the dominant cyclical factor that captures the common movements in credit growth. Compared to the baseline model, the second-order component is much more volatile, with the h_{xx} parameter estimated at -2.13 compared to -0.13 (Table 8 in Appendix Section H). This volatility can also be seen from the dashed black line in Figure 5, which shows the smoothed estimate of the second-order component. In contrast, the solid gray line, which is the first-order component, loses much of its cyclicality and instead plays a secondary role in capturing deviations of nonfinancial business and household credit from their movements generated by the second-order component. The factor loading estimates also reflect this result. The second-order component loads on all four credit growth series whereas the first-order component primarily loads onto nonfinancial business and household credit growth. The first-order component is still crucial, however, as it is the driver of fluctuations in the second-order component.

[Figure 5 about here]

The second takeaway is that the nonlinearities in the second-order component from this model are much stronger than those in the baseline. This comes from the larger estimate of h_{xx}, which

is the dependence of the second-order component on the squared first-order component. Figure 6, which shows the IRFs of the first- and second-order components in three historical episodes with differing conditions, and Figure 7, which shows the mean, standard deviation, and tail risk responses of the second-order component during the credit crunch, support this notion. The first figure illustrates the state-dependence of the second-order component, where the effects of shocks in the credit boom period are much milder than in the credit crunch or credit mix period. There also is evidence of an asymmetric response, where shocks that lead to declines in the second-order component have larger effects than those that lead to increases. The second figure shows that during the credit crunch, both the mean and volatility of the second-order component have powerful responses and move in opposite directions. The combination of these effects means that the shortfall of the second-order component is much more volatile compared to the longrise.

[Figure 6 about here]

[Figure 7 about here]

Finally, and perhaps most importantly, separating the nonlinear factor into its first- and second-order components reveals a state dependence that is obscured when the nonlinear factor is analyzed as a single object. In particular, the response of the second-order component to shocks can change sign across states. As shown in the first row of Figure 6, a positive shock lowers the second-order component during the credit crunch episode, but raises it during the credit mix period. This asymmetry depends on the sign of the first-order component at the time of the shock. When the first-order component is positive – as during the credit crunch – an increase in the first-order component raises its squared value, which reduces the second-order component given that $h_{xx} < 0$. Conversely, when the first-order component is negative – as during the credit mix – an increase in the first-order component lowers its squared value and therefore raises the second-order component. The effects of a negative shock follow the same logic in reverse.

This mechanism also shapes the higher-order moments of the second-order factor. Figure 7 shows that during the credit crunch, a positive shock reduces the mean of the second-order component while increasing its volatility, whereas a negative shock has the opposite effect. One interpretation is that the second-order factor captures the nonlinear dynamics of a credit boom turning

into a bust. When the first-order factor is negative or modest in magnitude, positive shocks increase the second-order component or have only limited effects – consistent with the standard view that credit expansions stimulate credit growth. However, when the first-order factor becomes sufficiently large, as during the GFC, positive shocks instead trigger a sharp decline in the second-order component accompanied by a surge in volatility: the credit boom has gone bust.

4.5 Further Robustness

In Appendix Section H, we show several additional important results. First, we compare our results from the NLDF model with those from a Markov-Switching Dynamic Factor Model (MSDF). The nonlinearities from the MSDF model are mainly due to state switches, which discretely change the mean and volatility of the factor. The MSDF model does not capture the smoothly-varying state dependence of the factor dynamics nor the hump-shaped responses of the factor to shocks during the credit crunch that the NLDF model generates. Second, we examine the robustness of our results to measurement errors with AR(1) persistence. We find that although the measurement errors for some series indeed have persistence, the estimates of the factor are similar across the models and the nonlinearities generated from the benchmark model survive.

Third, we estimate a two factor version of the NLDF model. We find that while the nonlinearities from the first factor become amplified, the second factor is approximately linear. Finally, we estimate an alternative model featuring a single factor with linear AR(1) dynamics and a quadratic measurement equation. The properties of this model are discussed in detail in Appendix Section G. It estimates a different credit cycle factor compared to the one from the benchmark NLDF model that is more consistent with the first-order factor in the NLDF model with separate weights on the first- and second-order factors. Finally, we show that our model can be easily modified to applications that require a nonlinear measurement equation. Specifically, we extract a common factor from a series of interest rates that respect the effective zero lower bound (see Appendix Section I).

5 Conclusion

We propose a parsimonious NLDF model that is built around a *pruned* second-order factor equation. In this model, the propagation of shocks is asymmetric, state dependent, and size dependent, and stationarity is easily verified. The application of the particle filter to evaluate the likelihood and extract the factor allows us consider environments in which the model dynamics are nonlinear. We use the model to examine the importance of nonlinearities in the U.S. credit cycle. This nonlinearity leads to state-dependent IRFs and changes in the higher-order moments in response to shocks.

Our work can be expanded in several directions. We mentioned already the multidimensional factor in the main text. Another fruitful avenue is to use the NLDF model with a VAR in the same fashion as the factor-augmented vector autoregression model (Stock and Watson, 2016).

References

T. Adrian, D. Covitz, and N. Liang. Financial Stability Monitoring. *Annual Review of Financial Economics*, 7(1):357–395, December 2015.

T. Adrian, N. Boyarchenko, and D. Giannone. Vulnerable Growth. *American Economic Review*, 109 (4):1263–1289, April 2019.

M. M. Andreasen, J. Fernandez-Villaverde, and J. F. Rubio-Ramirez. The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications. *The Review of Economic Studies*, 85(1):1–49, 06 2017.

B. Aruoba, L. Bocola, and F. Schorfheide. Assessing dsge model nonlinearities. *Journal of Economic Dynamics and Control*, 83:34–54, 2017.

S. B. Aruoba, M. Mlikota, F. Schorfheide, and S. Villalvazo. Svars with occasionally-binding constraints. *Journal of Econometrics*, 2021. ISSN 0304-4076.

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, October 2008.

J. Bai and P. Wang. Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics*, 33(2):221–240, April 2015. doi: 10.1080/07350015.2014.941467. URL https://ideas.repec.org/a/taf/jnlbes/v33y2015i2p221-240.html.

S. R. Baker, N. Bloom, and S. J. Davis. Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016. doi: 10.1093/qje/qjw024.

M. Banbura and M. Modugno. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160, 2014.

G. Benigno, P. Benigno, and S. Nistico. Second-order approximation of dynamic models with time-varying risk. *Journal of Economic Dynamics and Control*, 37(7):1231–1247, 2013.

B. S. Bernanke, M. Gertler, and S. Gilchrist. The financial accelerator in a quantitative business cycle framework. In J. B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 21, pages 1341–1393. Elsevier, December 1999.

N. Bloom. The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685, May 2009.

L. Bocola. The pass-through of sovereign risk. *Journal of Political Economy*, 124(4):879–926, 2016.

J.-J. Brady, B. Cox, Y. Li, and V. Elvira. Pydpf: A python package for differentiable particle filtering, 2025. URL https://arxiv.org/abs/2510.25693.

M. K. Brunnermeier and Y. Sannikov. A Macroeconomic Model with a Financial Sector. *American Economic Review*, 104(2):379–421, February 2014.

M. Chauvet. An econometric characterization of business cycles dynamics with factor structure and regime switching. *International Economic Review*, 39(4):969–996, 1998.

M. Chen, I. Fernández-Val, and M. Weidner. Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2):296 – 324, 2021. ISSN 0304-4076. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

X. Cheng, Z. Liao, and F. Schorfheide. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies*, 83:1151–1543, 2016.

M. Del Negro and C. Otrok. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York, 2008.

M. Deng, P. A. Guerrón-Quintana, and L. Tseng. Parallel computation of sovereign default models. *Computational Economics*, 62:1047–1085, 2023. doi: 10.1007/s10614-022-10291-1.

J. Fernández-Villaverde and J. F. Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087, 2007.

J. Fernández-Villaverde and J. F. Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087, 2007. doi: 10.1111/j.1467-937X.2007.00437.x.

J. Fernandez-Villaverde, G. Gordon, P. Guerron-Quintana, and J. F. Rubio-Ramirez. Nonlinear adventures at the zero lower bound. *Journal of Economic Dynamics and Control*, 57:182–204, 2015. ISSN 0165-1889.

J. Fernández-Villaverde, P. Guerrón-Quintana, K. Kuester, and J. Rubio-Ramírez. Fiscal volatility shocks and economic activity. *American Economic Review*, 105(11):3352–84, 2015a.

J. Fernández-Villaverde, P. Guerrón-Quintana, and J. F. Rubio-Ramírez. Estimating dynamic equilibrium models with stochastic volatility. *Journal of Econometrics*, 185(1):216–229, 2015b.

J. Fernandez-Villaverde, J. Rubio-Ramirez, and F. Schorfheide. Solution and estimation methods for dsge models. In J. Taylor and H. Uhlig, editors, *Handbook of Macroeconomics*, chapter 9, pages 557–724. North-Holland, 2016.

M. Gertler and P. Karadi. A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34, January 2011.

J. Geweke. Using simulation methods for bayesian econometric models: inference, development, and communication. *Econometric Reviews*, 18(1):1–73, 1999.

S. Gilchrist and E. Zakrajsek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, June 2012.

J. Gomes, U. Jermann, and L. Schmid. Sticky Leverage. *American Economic Review*, 106(12):3800–3828, December 2016.

S. Goncalves, A. M. Herrera, L. Kilian, and E. Pesavento. Impulse response analysis for structural dynamic models with nonlinear regressors. *Journal of Econometrics*, 2021. ISSN 0304-4076.

Y. Gorodnichenko and S. Ng. Level and volatility factors in macroeconomic data. *Journal of Monetary Economics*, 91:52–68, 2017.

C. Gourieroux, A. Monfort, and J.-P. Renne. Identification and Estimation in Non-Fundamental Structural VARMA Models. *The Review of Economic Studies*, 05 2019.

L. Guerrieri and M. Iacoviello. Collateral constraints and macroeconomic asymmetries. *Journal of Monetary Economics*, 90(C):28–49, 2017.

R. S. Gurkaynak, B. Sack, and J. H. Wright. The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304, November 2007.

C. Gust, E. Herbst, D. López-Salido, and M. E. Smith. The empirical implications of the interest-rate lower bound. *American Economic Review*, 107(7):1971–2006, July 2017a. doi: 10.1257/aer.20121437. URL https://www.aeaweb.org/articles?id=10.1257/aer.20121437.

C. Gust, E. Herbst, D. López-Salido, and M. E. Smith. The empirical implications of the interest-rate lower bound. *American Economic Review*, 107(7):1971–2006, 2017b.

E. Herbst and F. Schorfheide. Tempered particle filtering. *Journal of Econometrics*, 210(1):26–44, 2019.

J. T. Hwang. Multiplicative errors-in-variables models with applications to recent data released by the u.s. department of energy. *Journal of the American Statistical Association*, 81(395):680–688, 1986.

O. Jorda, M. Schularick, and A. M. Taylor. Sovereigns Versus Banks: Credit, Crises, And Consequences. *Journal of the European Economic Association*, 14(1):45–79, February 2016.

K. Jurado, S. C. Ludvigson, and S. Ng. Measuring uncertainty. *American Economic Review*, 105(3):1177–1216, March 2015. doi: 10.1257/aer.20131193.

A. Justiniano and G. E. Primiceri. The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3):604–41, 2008.

C.-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22, 1994.

D. H. Kim and K. J. Singleton. Term structure models and the zero bound: An empirical investigation of Japanese yields. *Journal of Econometrics*, 170(1):32–49, 2012.

J. Kim, S. Kim, E. Schaumburg, and C. A. Sims. Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, 32(11):3397–3414, 2008. ISSN 0165-1889.

D. Leiva-León, G. Perez-Quiros, H. Sapriza, and E. Zakrajšek. Introducing the Credit Market Sentiment Index. *Richmond Fed Economic Brief*, 22(33), August 2022.

F. Lindsten, M. I. Jordan, and T. B. Schön. Particle gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15(63):2145–2184, 2014.

S. C. Ludvigson and S. Ng. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222, January 2007.

S. C. Ludvigson, S. Ma, and S. Ng. Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410, October 2021.

J. Magnus and H. Neudecker. The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic and Discrete Methods*, 1(4):422–449, 1980. ISSN 0196-5212. Pagination: 28.

A. Mian, A. Sufi, and E. Verner. Household Debt and Business Cycles Worldwide. *The Quarterly Journal of Economics*, 132(4):1755–1817, 2017.

A. Onatski and C. Wang. Spurious Factor Analysis. *Econometrica*, 89(2):591–614, March 2021.

S. Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

M. Schularick and A. M. Taylor. Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008. *American Economic Review*, 102(2):1029–1061, April 2012.

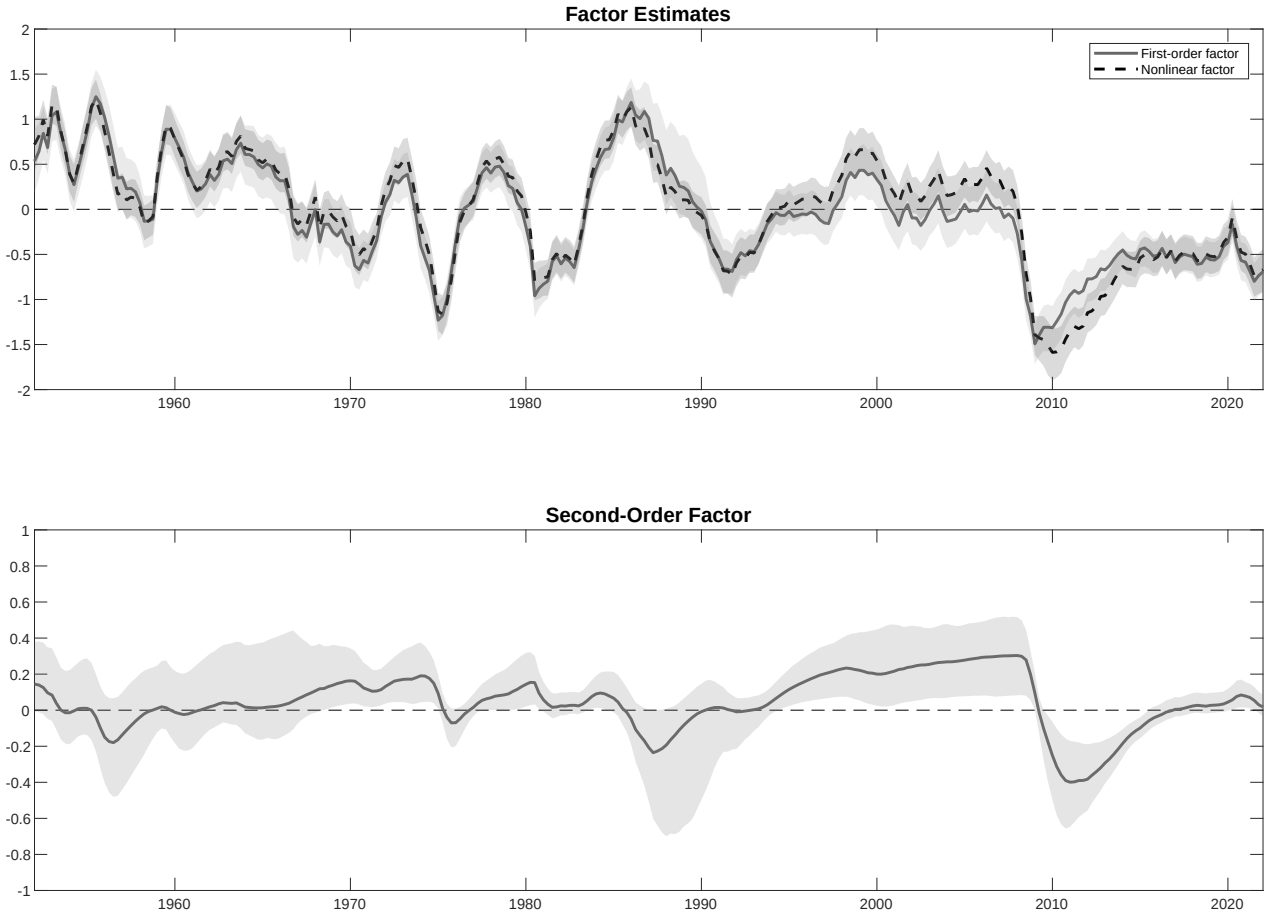
M. Shintani. Nonlinear forecasting analysis using diffusion indexes: An application to japan. *Journal of Money, Credit, and Banking*, 37(3):517–538, 2005.

J. Stock and M. Watson. Factor models and structural vector autoregressions in macroeconomics. In J. Taylor and H. Uhlig, editors, *Handbook of Macroeconomics*, chapter 8, pages 415–526. North-Holland, 2016.

J. C. Wu and F. D. Xia. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291, March 2016.

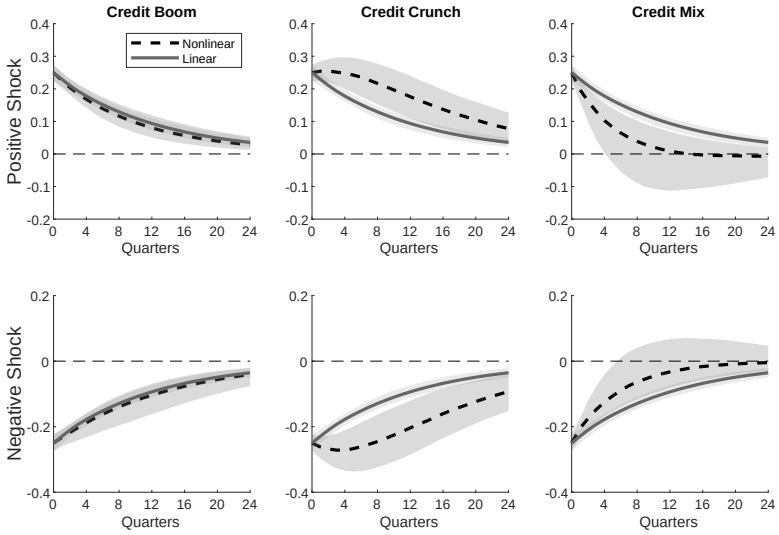
6 Supplemental materials

Figure 1: Estimated Credit Cycle and the Contribution of the Second-Order Factor



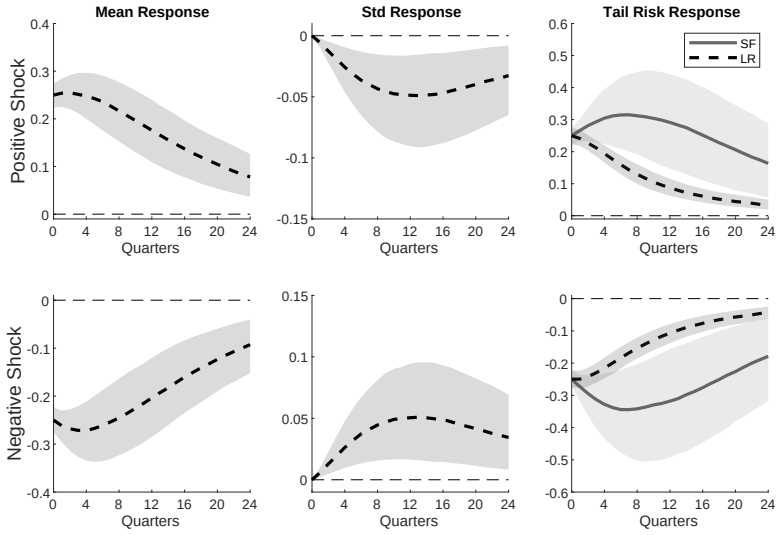
NOTE: Smoothed factor estimates produced by the nonlinear dynamic factor model. The top panel shows the demeaned factor estimates, with the dashed line being the estimate of the nonlinear factor and the solid line the estimate of the first-order factor. The shaded areas denote 68% credible sets. The bottom panel shows the demeaned second-order factor with 68% credible sets.

Figure 2: State-Dependent Impulse Response Functions in Three Periods



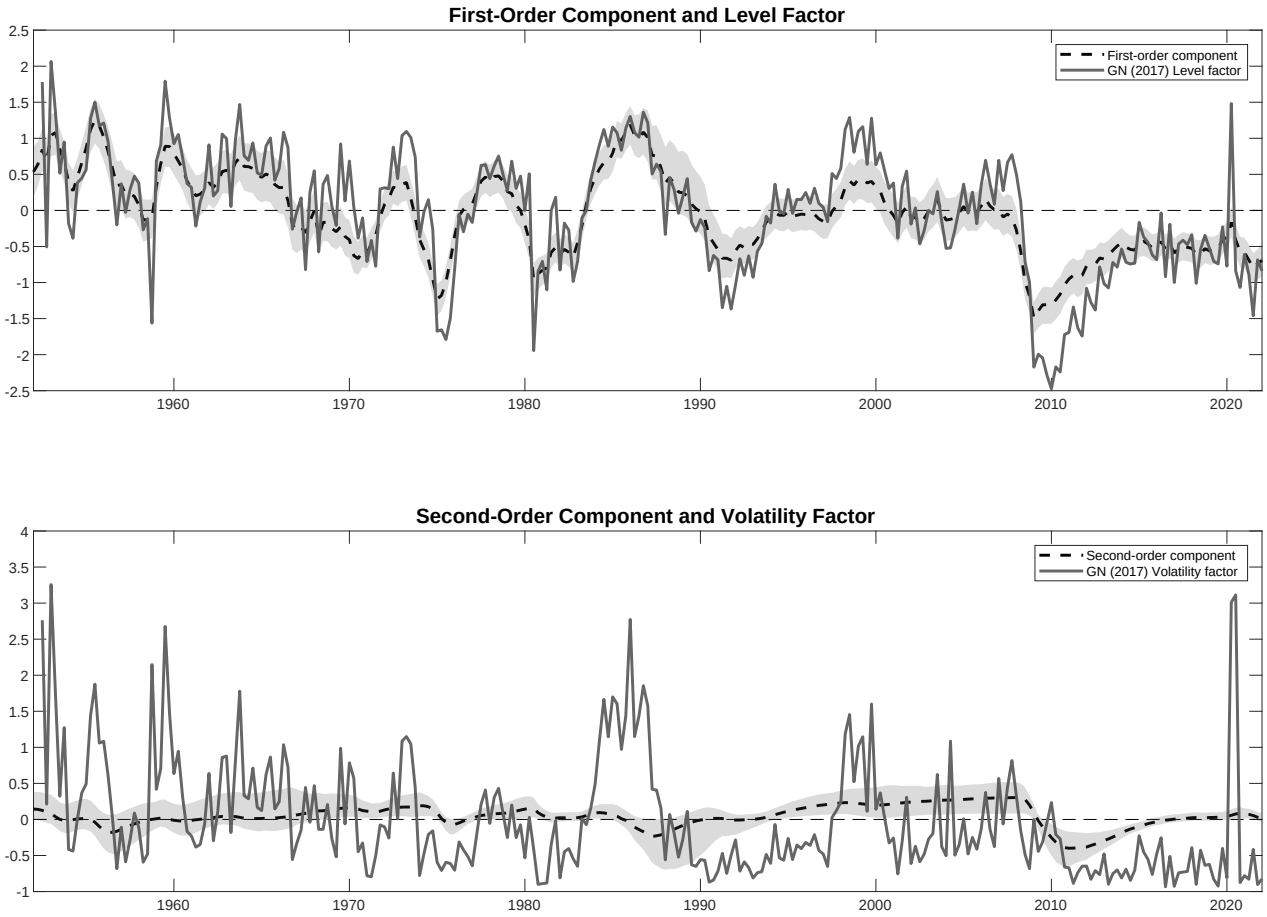
NOTE: The dashed lines denote the responses of the overall factor following a positive shock (top row) and negative shock (bottom row), while the solid lines denote the responses of the linear component of the model. The first column conditions on a credit boom period in the mid-2000s, the second column conditions on a credit bust period in 2010, and the third column conditions on a mixed case before the early 1990s recession. The shaded areas denote 68% credible sets.

Figure 3: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk During the Credit Crunch



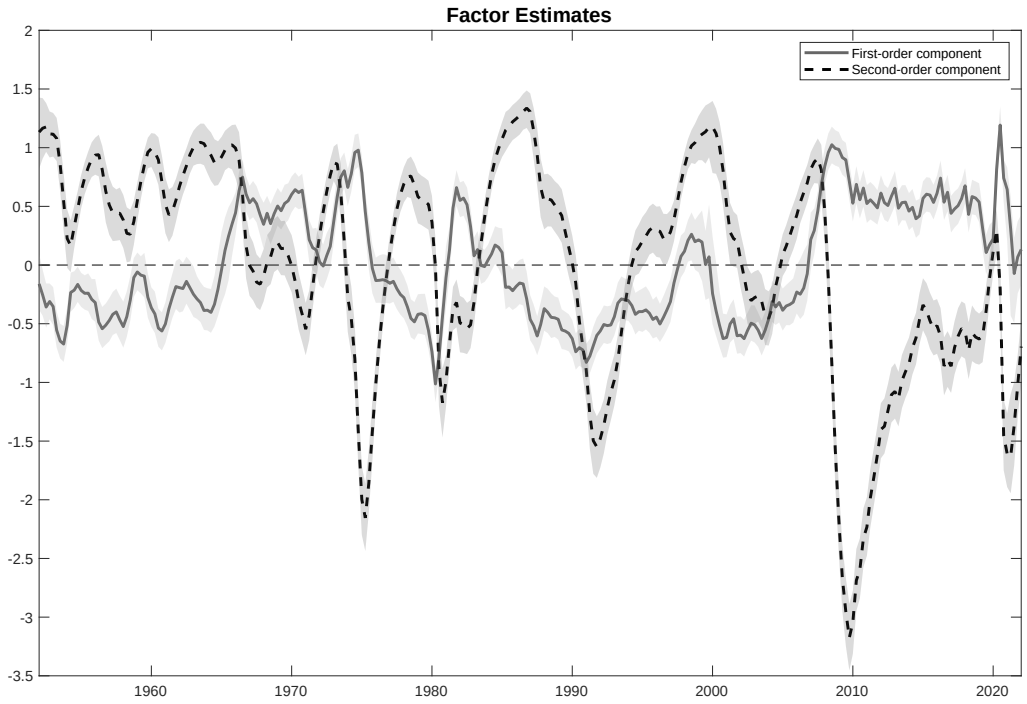
NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the demeaned overall factor produced by the nonlinear dynamic factor model during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. In the third column, the solid lines denote the shortfall (SF) response, while the dashed lines denote the longrise (LR) response. The shaded areas are 68% credible sets.

Figure 4: Comparing NLDF and Gorodnichenko and Ng (2017) Estimates



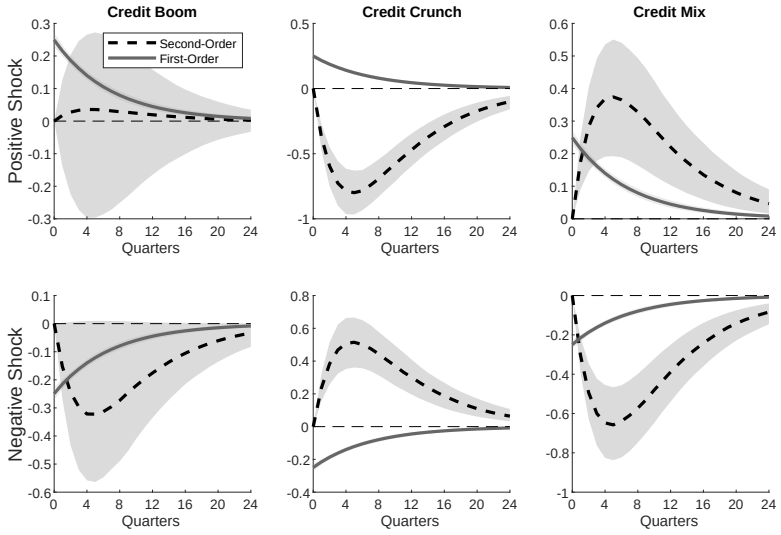
NOTE: Smoothed first- and second-order component estimates (dashed lines) produced by the nonlinear dynamic factor model compared to the linear and volatility factors produced using the methodology of Gorodnichenko and Ng (2017) (solid lines). The top panel shows the first-order component and level factor and the bottom panel shows the second-order component and volatility factor. The shaded areas denote 68% credible sets.

Figure 5: Estimated First- and Second-Order Components



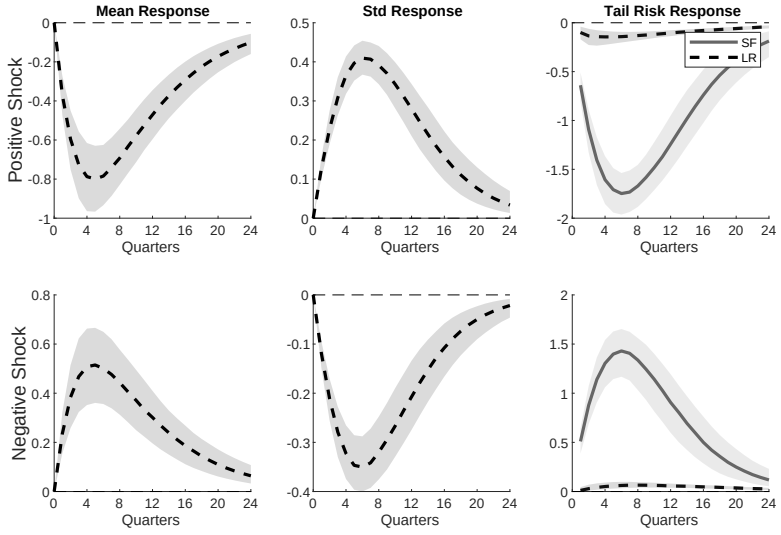
NOTE: Smoothed first- and second-order component estimates produced by the nonlinear dynamic factor model. The solid line is the estimate of the first-order component and the dashed line the estimate of the second-order component. The shaded areas denote 68% credible sets.

Figure 6: Extended Model: State-Dependent Impulse Response Functions in Three Periods



NOTE: The dashed lines denote the responses of the second-order component following a positive shock (top row) and negative shock (bottom row), while the solid lines denote the responses of the first-order component. The first column conditions on a credit boom period in the mid-2000s, the second column conditions on a credit bust period in 2010, and the third column conditions on a mixed case before the early 1990s recession. The shaded areas denote 68% credible sets. Note that since we have different factor loadings for the first- and second-order components, we cannot show an overall nonlinear factor as before.

Figure 7: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk of the Second-Order Factor During the Credit Crunch



NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the demeaned second-order component produced by the model during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. In the third column, the solid lines denote the shortfall (SF) response, while the dashed lines denote the longrise (LR) response. The shaded areas are 68% credible sets.

NOT FOR PUBLICATION

Appendices

A Identification

In this section, we discuss factor identification issues in our model. First, we consider two cases with linear measurement equations. Second, we move onto identification in the interest rate application, which has a nonlinear measurement equation.

A.1 Linear Measurement Equation

We consider the specification with a linear measurement equation. y_t is a $N \times 1$ vector of observables and ϵ_t is a $N \times 1$ vector of measurement error innovations. f_t is a $K \times 1$ vector of factors and ν_t is a $K \times 1$ vector of factor innovations. G is a $N \times K$ matrix, η is an $N \times N$ diagonal matrix of standard deviations, h_x is a $K \times K$ matrix, σ is a $K \times K$ matrix square root of the covariance matrix, and h_{xx} is a $K \times \tilde{K}$ matrix, where $\tilde{K}$ denotes the number of unique elements in the Kronecker product $f_{t-1}^f \otimes f_{t-1}^f$.

Case 1 (Proposition 2.1): Common Loadings on f_t^f and f_t^s

$$
\begin{aligned}
&y_t = Gf_t + \eta\epsilon_t \\
&\begin{cases}
f_t = c + f_t^f + f_t^s \text{ – first- and second-order components} \\
f_t^f = h_x f_{t-1}^f + \sigma\nu_t \\
f_t^s = h_x f_{t-1}^s + \frac{1}{2} h_{xx} vech\left(f_{t-1}^f \otimes f_{t-1}^f\right).
\end{cases}
\end{aligned}
$$

This model can equivalently be written as:

$$
\begin{aligned}
&y_t = Gf_t + \eta\epsilon_t \\
&\begin{cases}
f_t = c + f_t^f + f_t^s \text{ – first- and second-order components} \\
f_t^f = h_x f_{t-1}^f + \sigma\nu_t \\
f_t^s = h_x f_{t-1}^s + \frac{1}{2}\underbrace{h_{xx}E}_{=\ddot{h}_{xx}}\left(f_{t-1}^f \otimes f_{t-1}^f\right).
\end{cases}
\end{aligned}
$$

where E denotes a $\tilde{K} \times K^2$ elimination matrix which selects the unique elements of $f_{t-1}^f \otimes f_{t-1}^f$ and $\ddot{h}_{xx} = h_{xx}E$. This representation allows for an interaction of the first-order components. In the case of 2 nonlinear factors: $f_{1,t-1}^f \times f_{2,t-1}^f$.[1]

Consider a full rank matrix A. Call $\tilde{f}_t^f = Af_t^f$, $\tilde{f}_t^s = Af_t^s$, and $\tilde{f}_t = Af_t$.

$$
\begin{aligned}
&y_t = GA^{-1}\tilde{f}_t + \eta\epsilon_t \\
&\begin{cases}
\tilde{f}_t = Ac + \tilde{f}_t^f + \tilde{f}_t^s \\
\tilde{f}_t^f = Ah_x A^{-1}\tilde{f}_{t-1}^f + A\sigma\nu_t \\
\tilde{f}_t^s = Ah_x A^{-1}\tilde{f}_{t-1}^s + \frac{1}{2}A\ddot{h}_{xx}\left(A^{-1} \otimes A^{-1}\right)\left(\tilde{f}_{t-1}^f \otimes \tilde{f}_{t-1}^f\right).
\end{cases}
\end{aligned}
$$

These equations can be rewritten as:

$$
\begin{aligned}
&y_t = \tilde{G}\tilde{f}_t + \eta\epsilon_t \\
&\begin{cases}
\tilde{f}_t = \tilde{c} + \tilde{f}_t^f + \tilde{f}_t^s \\
\tilde{f}_t^f = \tilde{h}_x \tilde{f}_{t-1}^f + \tilde{\sigma}\nu_t \\
\tilde{f}_t^s = \tilde{h}_x \tilde{f}_{t-1}^s + \frac{1}{2}\tilde{\ddot{h}}_{xx}\left(\tilde{f}_{t-1}^f \otimes \tilde{f}_{t-1}^f\right).
\end{cases}
\end{aligned}
$$

where $\tilde{G} = GA^{-1}, \tilde{c} = Ac, \tilde{h}_x = Ah_xA^{-1}$, and $\tilde{\ddot{h}}_{xx} = A\ddot{h}_{xx}\left(A^{-1} \otimes A^{-1}\right)$. Moreover, $\tilde{\ddot{h}}_{xx}$ can be

[1] For example, in the two factor case, $E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

further decomposed as $\tilde{h}_{xx}E$, where $\tilde{h}_{xx} = A\ddot{h}_{xx}\left(A^{-1} \otimes A^{-1}\right) D$ and D is a $K^2 \times \tilde{K}$ duplication matrix.[2] The last decomposition relies on the result that $DE\left(\tilde{f}^f_{t-1} \otimes \tilde{f}^f_{t-1}\right) = \left(\tilde{f}^f_{t-1} \otimes \tilde{f}^f_{t-1}\right)$ since $\left(\tilde{f}^f_{t-1} \otimes \tilde{f}^f_{t-1}\right) = vec\left(\tilde{f}_{t-1}\tilde{f}'_{t-1}\right)$ and $\tilde{f}_{t-1}\tilde{f}'_{t-1}$ is a symmetric matrix.[3]

This means that we have the usual rotation and scale indeterminacy. Moreover, once we impose the usual identification conditions for linear dynamic factor models (Bai and Wang, 2015), our model is also identified. We can continue to use this identification strategy because the first- and second-order components enter additively with the same factor loadings into the measurement equation. Therefore, the presence of the second-order components does not affect the identifiability of the first-order components. Since the second-order components are functions of past first-order components and they load onto the measurement equation with the same factor loadings, once we identify the first-order components, the second-order components are also identified.

Case 2 (Proposition 2.2): Separate Weights on the First- and Second-Order Components

In this model, we allow for separate weights on the first- and second-order components of the factors in the measurement equation. In addition, we allow the autoregressive matrix $h_{x,1}$ of the first-order component to be different than that of the second-order component $h_{x,2}$.[4] The baseline NLDF model requires equal autoregressive matrices across first- and second-order components, but this restriction becomes unnecessary when extending the model to allow separate weights for each component.

f^f_t is a $K \times 1$ vector of first-order components and f^s_t is a $K \times 1$ vector of second-order components. G_1 and G_2 are $N \times K$ matrices. h^1_x and h^2_x are $K \times K$ matrices.

[2] For example, in the two factor case, $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

[3] See Magnus and Neudecker (1980).

[4] The assumption of separate persistence terms is natural given that we allow for separate effects of the first- and second-order factors on the observables. The second-order factor can be interpreted as closer to a separate factor as opposed to being tightly linked to the first-order factor. This assumption also makes the identification discussion more straightforward.

$$y_t = G_1 f_t^f + G_2 \left(c + f_t^s\right) + \eta \epsilon_t$$
$$\begin{cases} f_t^f = h_{x,1} f_{t-1}^f + \sigma \nu_t \\ f_t^s = h_{x,2} f_{t-1}^s + \frac{1}{2} h_{xx} vech\left(f_{t-1}^f \otimes f_{t-1}^f\right). \end{cases}$$

This model can equivalently be written as:

$$y_t = G_1 f_t^f + G_2 \left(c + f_t^s\right) + \eta \epsilon_t$$
$$\begin{cases} f_t^f = h_{x,1} f_{t-1}^f + \sigma \nu_t \\ f_t^s = h_{x,2} f_{t-1}^s + \frac{1}{2} \underbrace{h_{xx} E}_{=\ddot{h}_{xx}} \left(f_{t-1}^f \otimes f_{t-1}^f\right). \end{cases}$$

Consider full rank matrices A_1 and A_2. Call $\tilde{f}_t^f = A_1 f_t^f$ and $\tilde{f}_t^s = A_2 f_t^s$.

$$y_t = \begin{pmatrix} G_1 A_1^{-1} & G_2 A_2^{-1} \end{pmatrix} \begin{pmatrix} \tilde{f}_t^f \\ \\ A_2 c + \tilde{f}_t^s \end{pmatrix} + \eta \epsilon_t$$
$$\begin{cases} \tilde{f}_t^f = A_1 h_{x,1} A_1^{-1} \tilde{f}_{t-1}^f + A_1 \sigma \nu_t \\ \tilde{f}_t^s = A_2 h_{x,2} A_2^{-1} \tilde{f}_{t-1}^s + \frac{1}{2} A_2 \ddot{h}_{xx} \left(A_1^{-1} \otimes A_1^{-1}\right) \left(\tilde{f}_{t-1}^f \otimes \tilde{f}_{t-1}^f\right). \end{cases}$$

$$y_t = \tilde{G}_1 \tilde{f}_t^f + \tilde{G}_2 \left(\tilde{c} + \tilde{f}_t^s\right) + \eta \epsilon_t$$
$$\begin{cases} \tilde{f}_t^f = \tilde{h}_{x,1} \tilde{f}_{t-1}^f + \tilde{\sigma} \nu_t \\ \tilde{f}_t^s = \tilde{h}_{x,2} \tilde{f}_{t-1}^s + \frac{1}{2} \tilde{\ddot{h}}_{xx} \left(\tilde{f}_{t-1}^f \otimes \tilde{f}_{t-1}^f\right). \end{cases}$$

where $\tilde{G}_1 = G_1 A_1^{-1}, \tilde{G}_2 = G_2 A_2^{-1}, \tilde{c} = A_2 c, \tilde{h}_{x,1} = A_1 h_{x,1} A_1^{-1}, \tilde{h}_{x,2} = A_2 h_{x,2} A_2^{-1}$, and $\tilde{\ddot{h}}_{xx} = A_2 \ddot{h}_{xx} \left(A_1^{-1} \otimes A_1^{-1}\right)$. We can further decompose $\tilde{\ddot{h}}_{xx} = \tilde{h}_{xx} E$, where $\tilde{h}_{xx} = A_2 \ddot{h}_{xx} \left(A_1^{-1} \otimes A_1^{-1}\right) D$.

To identify the first-order components, we use the strategies in Bai and Wang (2015). The difference relative to Case 1 is that identifying the first-order components is insufficient to identify

the second-order components. This is because the second-order components load separately onto the measurement equation. Therefore, even when the first-order components are identified, there still is a matrix A_2 that can affect the identification of the second-order components. To resolve the identification problem for the second-order components, we additionally impose restriction DFM2 from Bai and Wang (2015): which in our context implies that a $K \times K$ block of G_2 is an identity matrix.

A.2 Measurement Equation in Interest Rate Application

We now also discuss identification in our interest rate application, which features a nonlinear measurement equation.

$$\Delta forward_t^h = m_h + \begin{cases} G_h(c + f_t^f + f_t^s) + \eta^h \epsilon_t^h & \text{if } \widehat{S}_t^h >= 0.3 \\ -m_h + \eta^h \epsilon_t^h & \text{otherwise} \end{cases}$$

where $c = -\frac{1}{2}\frac{h_{xx}\sigma^2}{(1-h_x)(1-h_x^2)}$, $\widehat{S}_t^h = \sum_{\tau=2}^{t}\left(m_h + G_h\left(c + f_\tau^f + f_\tau^s\right)\right) + forward_1^h$, $\Delta forward_t^h = forward_t^h - forward_{t-1}^h$, and index h stands for the maturity.

$$\begin{cases} f_t = c + f_t^f + f_t^s \text{ – first- and second-order components} \\ f_t^f = h_x f_{t-1}^f + \sigma \nu_t \\ f_t^s = h_x f_{t-1}^s + \frac{1}{2} h_{xx}\left(f_{t-1}^f\right)^2. \end{cases}$$

Another way of writing the measurement equation is:

$$\Delta forward_t = m + diag\left(\mathbb{1}\left(\widehat{S}_t >= 0.3\right)\right) G f_t + \left(I_N - diag(\mathbb{1}\left(\widehat{S}_t >= 0.3\right)\right)(-m) + \eta \epsilon_t$$

where $diag\left(\mathbb{1}\left(\widehat{S}_t >= 0.3\right)\right)$ is a diagonal matrix with a 1 if the corresponding element of $\widehat{S}_t >= 0.3$ and 0 otherwise. I_N is an $N \times N$ identity matrix.

$$\widehat{S}_t = \sum_{\tau=2}^{t} (m + Gf_t) + forward_1$$

Note that although the measurement equation is nonlinear, the factors always enter in the form Gf_t. Therefore, the identification issues and solutions discussed in Case 1 can be applied here as well.

B Variants to the Nonlinear Dynamic Factor Model

Beyond a Second-Order Representation

Our choice of the second-order pruned representation for the factor dynamics is based on its being parsimonious and on macroeconomists' familiarity with perturbation methods. But our exposition is general enough that one can use, for example, projection methods to approximate the functions $\mathcal{G}$ and $\mathcal{H}$. This alternative can capture richer nonlinearities that monomials cannot model. Let $\Psi_i(\cdot)$ denote the Chebyshev polynomial of degree i. Then the nonlinear state equation can be approximated by

$$f_t = \sum_{i=0}^{n} \theta_i \Psi_i(\widetilde{f}_{t-1}) + \sigma \nu_t.$$

Here, θ_i are parameters to be estimated and $\widetilde{f}_{t-1}$ is a transformation of the original $t-1$ factor such that it is bounded between -1 and 1.[5] However, this option comes at the cost of a potentially more complex likelihood when estimating the model.

Multidimensional State

One can, in theory, easily expand our model to accommodate more factors. Below, we still use two observables, but we add an additional factor.

$$\underset{2\times 1}{\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}} = \underset{2\times 2}{G} \underset{2\times 1}{\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}} + \underset{[2\times 2][2\times 1]}{\eta \; \epsilon_t},$$

$$\underset{2\times 1}{\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}} = \mathcal{H}\left(\underset{2\times 1}{\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix}} \right) + \underset{2\times 2}{\Sigma} \nu_t.$$

Here, the function $\mathcal{H}(\cdot)$ is the nonlinear map between the factors yesterday and the factors today. If one extends the pruned representation from above to the two-factor case, the results from Proposition 2.1 carry over. The identification strategies discussed in Appendix Section A.1 can be applied.

[5] This transformation is necessary because Chebyshev polynomials are defined in the interval $[-1, 1]$.

C Additional Analytical Properties of the Model

Unconditional Relationship Between Mean and Volatility

The unconditional variance-covariance matrix of the system is given by Equations 16 through 18:

$$V\left(z\right) = AV\left(z\right)A' + BV\left(\zeta\right)B' \tag{16}$$

$$V(\zeta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & E\left(f^f\right)^2 \end{pmatrix} \tag{17}$$

$$E\left(z\right) = \left(I - A\right)^{-1} \begin{pmatrix} 0 \\ 0 \\ \sigma^2 \end{pmatrix}. \tag{18}$$

The general form of $V(z)$ is as follows:

$$V(z) = \begin{pmatrix} X & 0 & 0 \\ 0 & X & X \\ 0 & X & X \end{pmatrix}, \tag{19}$$

where the X denotes nonzero values. Unconditionally, there is no correlation between the first-order factor and its square, or between the first- and second-order factors. Although the first-order factor determines the time-varying volatility in the system, it unconditionally has zero correlation with volatility because the time-varying volatility depends only on the magnitude of f^f_{t-1}, not its sign. As the unconditional distribution of f^f_t is symmetric around zero, this correlation is also zero unconditionally.

There is dependence, however, between the second-order factor and the volatility in the system, which induces a relationship between level and volatility, even unconditionally. This dependence

arises because the first-order factor squared enters as the driving force of the second-order factor. Unsurprisingly, the sign of h_{xx} is important in governing this relationship, with a positive h_{xx} generating a positive dependence between level and volatility and a negative h_{xx} generating a negative dependence.

D Simulation Results

We use a parameterized version of the nonlinear model and simulation methods to further illustrate its properties. To this end, let us consider the baseline model with two observables:

$$\begin{aligned}\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} &= \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} f_t + \eta \epsilon_t \\ f_t &= c + f_t^f + f_t^s \\ f_t^f &= h_x f_{t-1}^f + \sigma v_t \\ f_t^s &= h_x f_{t-1}^s + \frac{1}{2} h_{xx} \left(f_{t-1}^f \right)^2 .\end{aligned} \tag{20}$$

The baseline parameterization is $c = -\frac{1}{2}\frac{h_{xx}\sigma^2}{(1-h_x)(1-h_x^2)}$, $G_1 = 1$, $G_2 = 2$, $h_x = 0.45$, $h_{xx} = 0.5$, and $\sigma = 1$. The setting of c guarantees that the overall factor f has zero mean. We start our discussion with the IRFs implied by the model. Then, we move on to the distributional implications of the shocks. A key parameter in the analysis is h_{xx}. Its sign determines the direction of the asymmetry. We set $h_{xx} > 0$, but if h_{xx} were to be negative, then the asymmetries would be flipped.

Impulse Response Functions. We focus on the three novel properties that our factor model can deliver: asymmetric, state-dependent, and size-dependent shocks.[6]

We begin by discussing the asymmetric responses to shocks. The first row in Figure 8 shows the IRFs of the factor following a one standard deviation positive innovation (left panel) and a negative one (right panel), initializing the first-order factor f_{-1}^f at 0.56 and the second-order factor f_{-1}^s at its unconditional mean value. This calibration is illustrative, and we choose a nonzero lag of the first-order factor to showcase the state dependence in the IRFs. The figure plots the dynamics of the factor (blue line) and its first- (dashed black line) and second-order (solid black line)

[6] As our model is nonlinear and state dependent, we use the definition of conditional IRFs discussed by Goncalves et al. (2021): $IRF_{\delta,t-1}\left[f_{t+h}\right] = E\left[f_{t+h}\left(\delta\right) - f_{t+h}|\Omega_{t-1}\right]$, where $\Omega_{t-1} = \{f_{t-1}^f, f_{t-2}^f, ..., f_{t-1}^s, f_{t-2}^s, ...\}$, δ is the size of the innovation, f_{t+h} is the baseline value conditional on a path of shocks $\{\nu_t, \nu_{t+1}, ...\}$, and $f_{t+h}\left(\delta\right)$ is the counterfactual value conditional on the same path of shocks except with the addition of δ at time t $\{\nu_t + \delta, \nu_{t+1}, ...\}$.

components.[7] With this parameterization and initial condition, a positive shock persists for longer than a negative shock. This divergence can be seen by comparing the impulse response with its first-order component. The first-order component is linear and therefore symmetric. It produces impulse responses that are representative of those that come from a standard DFM. The blue line is formed by adding the first- and second-order components together. Given the initial conditions and the fact that $h_{xx} > 0$, the second-order term is always positive in this example. This finding means that the blue line is always above the first-order response, no matter whether the shock is positive or negative.

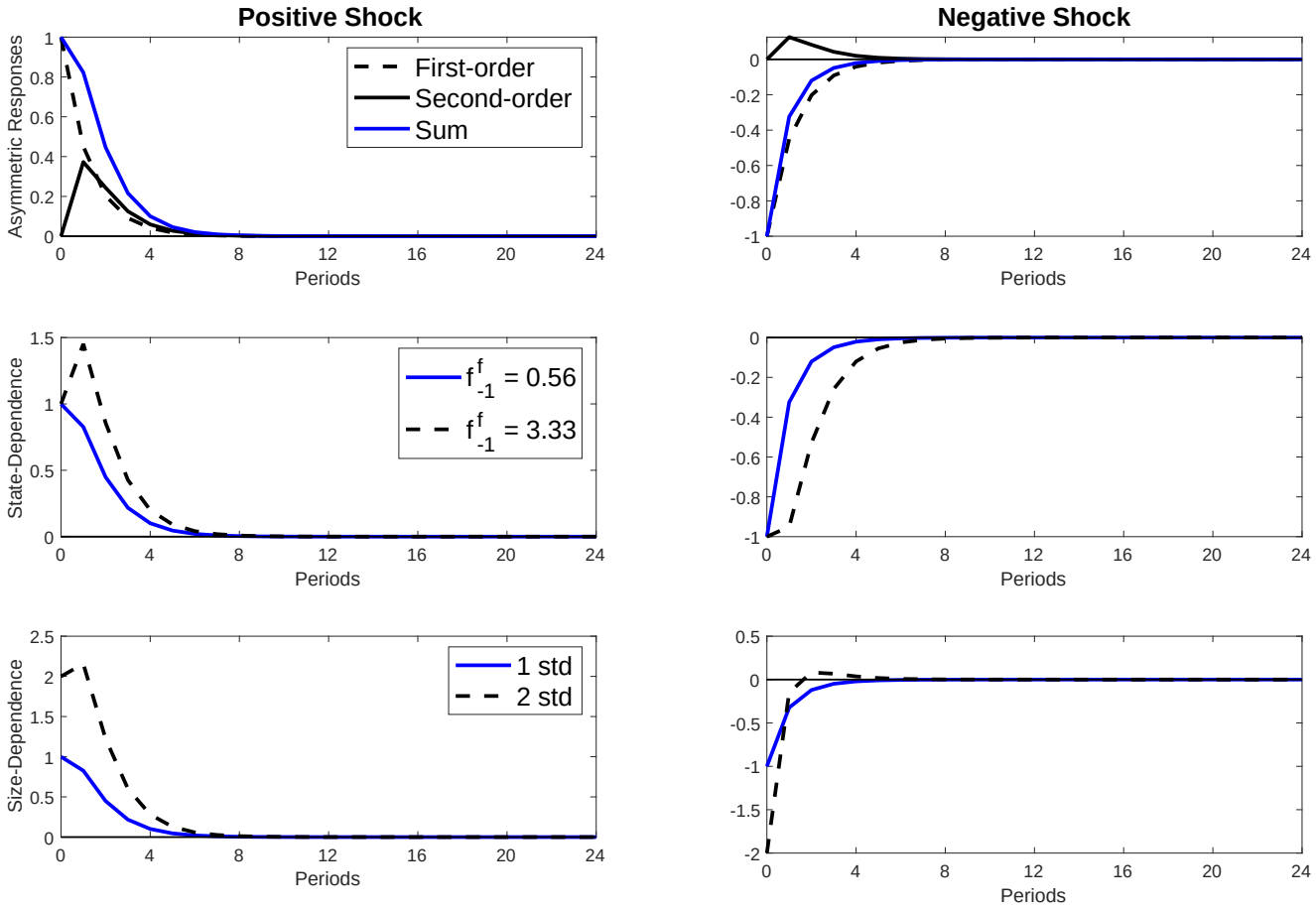
The second row of the figure illustrates the next important property that our model can produce: state dependence. The solid blue line is the same response as in the top row. The dashed black line now shows the responses to the same-sized shock but starting at an initial condition of $f^f_{-1} = 3.33$ and f^s_{-1} at its unconditional mean. Although the initial impulse is the same, the effects of the state dependence kick in with a one period lag. Starting from the different initial condition with an elevated first-order factor, both positive and negative shocks generate larger magnitude of responses in the factor. This difference can be understood by examining Equation 10. The lag of the first-order factor determines the volatility of ϵ_t in the third equation governing $\left(f^f_t\right)^2$, with a f^f_{t-1} that is larger in magnitude leading to a higher variance of ϵ_t.

The responses shown in the second row of the figure lead us to another related fact, which we illustrate in more detail in the section on distributional responses to shocks but is worth mentioning here. The different amplification of shocks is indicative of time-varying volatility. This example illustrates that when the first-order component starts out at a larger value in magnitude, the overall nonlinear factor also becomes more volatile. We contrast this state dependence with factor models of exogenous stochastic volatility, such as Del Negro and Otrok (2008), where movements in the volatility of the factor are due to separate shocks.

Finally, the third row of Figure 8 shows the size dependence of the IRFs. What we mean by size dependence is that the shape of the IRF changes depending on the size of the shock. The

[7] We generate the distributional responses and IRFs by simulating $100,000$ paths from the initial condition. In the baseline case, we take draws from the data generating process. In the Positive Shock case, we add a one standard deviation shock to the impact period's innovations from the baseline. After the impact period, we use the exact same draws of the innovations in both scenarios. The IRFs are differences in the conditional moments of the resulting distributions.

Figure 8: Nonlinear Dynamic Factor Model Can Generate Asymmetric, State-Dependent, and Size-Dependent Responses to Shocks



NOTE: **Row 1**: Asymmetric responses to the same-sized shock in a calibrated model. This figure shows the effects of a positive shock on the left panel in blue. The dashed black line is the first-order response, while the solid black line is the second-order response. **Row 2**: State-dependent responses in a calibrated model. The responses to a positive shock are shown in the left panel. The blue line is the same as in Row 1, and the dashed black line is the response to the same shock but at a different initial condition. **Row 3**: Size-dependent responses in a calibrated model. The responses to a positive shock are shown in the left panel. The blue line is the same as in Row 1, and the dashed black line is the response to a shock twice the size as the one that generates the blue line. In all rows, the right panel shows the responses to a negative shock.

blue line is the same one that we have carried over from the previous rows. The dashed black line is the response at the same initial conditions but to a shock that is two standard deviations in size instead of one standard deviation. On impact, the response is double that of the one standard deviation shock. In the next period, however, the shapes of the IRF change for positive and negative shocks. This fact is especially clear in this example after a negative shock, in which the IRF turns

positive two periods after impact following a two standard deviation shock, while it stays negative following a one standard deviation shock. In essence, a two standard deviation change to the first-order component changes the second-order component by more than double that of the first-order component. This greater-than-proportionate response of the second-order component generates the different shape of the overall IRF.

Distributional Implications Despite being driven by normal, homoskedastic shocks, the NLDF model produces rich non-normalities in the distribution of the factor, which then feeds into the distribution of the observables. We begin by discussing the unconditional distribution of the factor. Next, we move on to the time-varying volatility and tail risk that our model can produce.

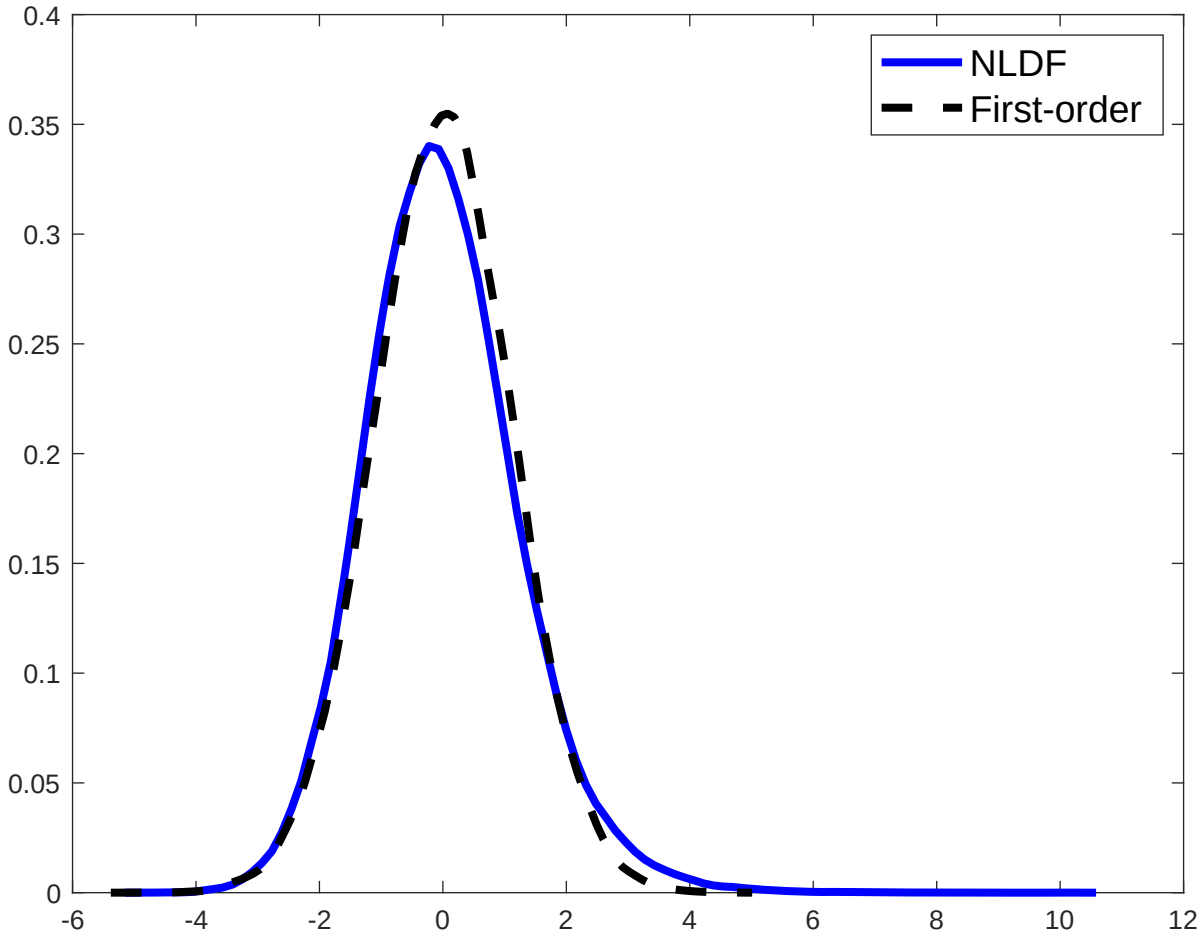
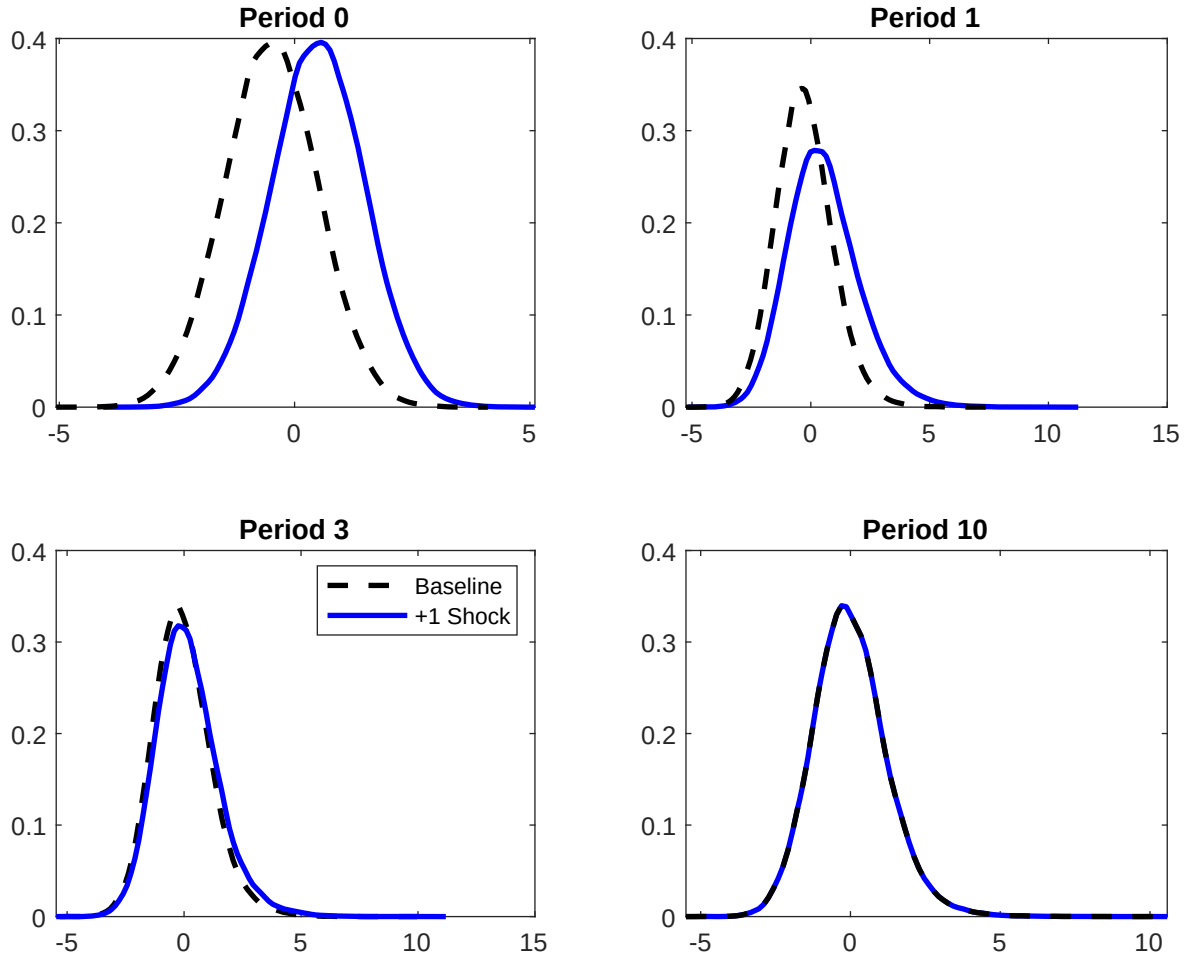


Figure 9: Unconditional distribution of the overall factor in a calibrated model. The blue line denotes the unconditional distribution of the demeaned overall factor in the nonlinear dynamic factor model. The dashed black line denotes the unconditional distribution of the first-order factor.

Figure 9 shows the unconditional distribution from a long simulation. The unconditional distribution produced by the NLDF model is not normal, as evidenced by its positive skew. Its Kelley skewness, which measures the share of the distance from the 90th percentile to the 10th percentile that is above the median versus below the median, is 0.07.[8] By contrast, the distribution produced by the first-order component is normally distributed and therefore has a Kelley skewness of 0.

Figure 10: Dynamics of the Distribution of the Demeaned Overall Factor in Response to a Positive Shock



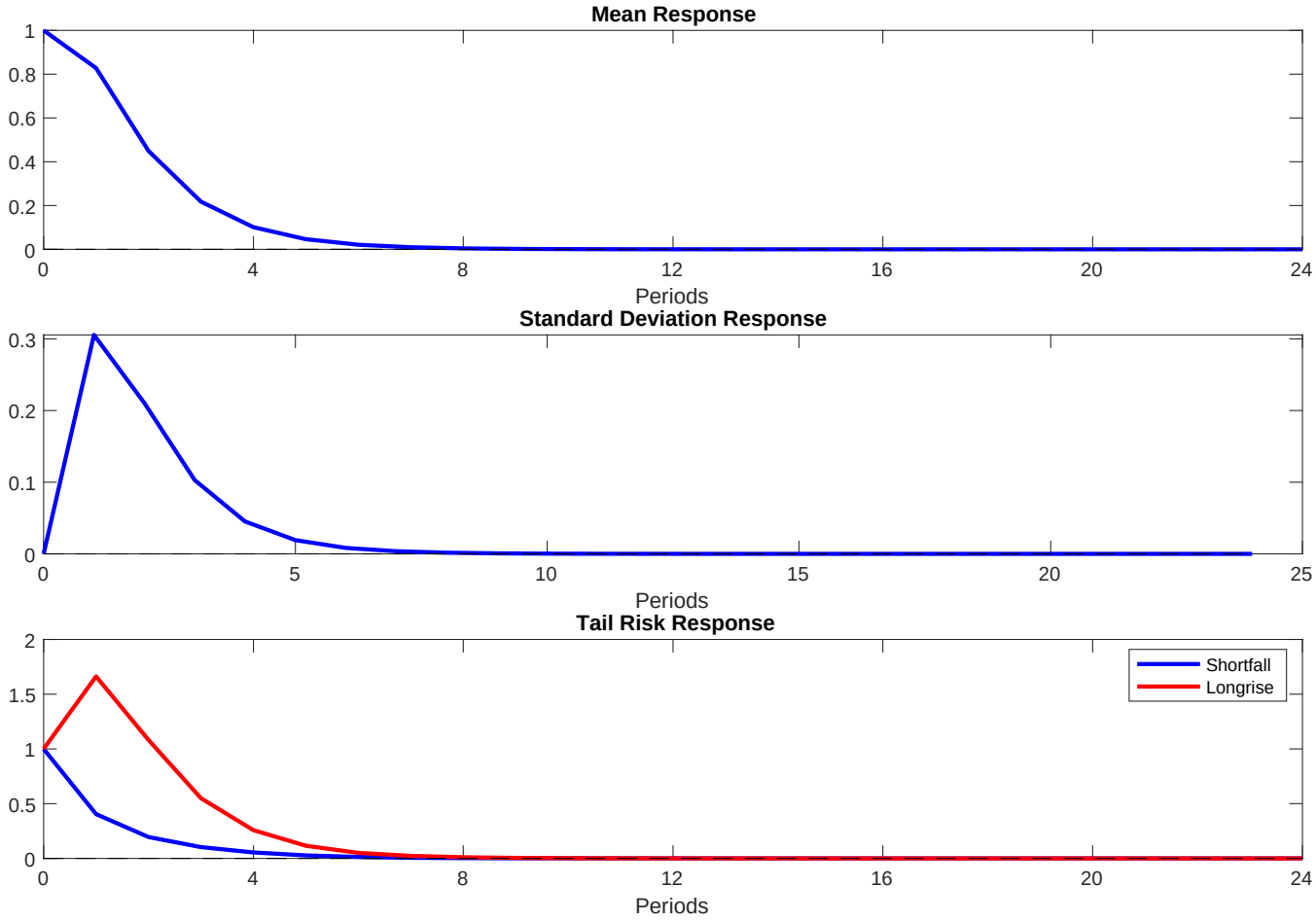
NOTE: The blue line is the distribution after a positive shock, and the dashed black line is the distribution without a shock. Period 0 is the period of the shock.

In Figure 10, we illustrate the distributional implications of shocks in our model. We do so by simulating the predictive distributions at various horizons following the same one standard

[8]The formula for Kelley skewness is $\frac{Q90+Q10-2*Q50}{Q90-Q10}$, where Q is the quantile of the distribution.

deviation shock that we began discussing in the top row of Figure 8. The blue distribution is the baseline density – the one that characterizes the possible outcomes if we simulate the model from the initial conditions. The dashed black line is the density that realizes if we had a positive one standard deviation shock at period 0. Our model implies that such a shock leads to a positive shift in the distribution on impact. Crucially, in period 1, the distribution *widens out*. The differences in the distributions persist through period 3, and by period 10, the effects of the shock are largely gone.

Figure 11: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk of the Demeaned Overall Factor in Response to a Positive Shock



NOTE: In the third panel, the blue line shows the response of the 5% shortfall, and the red line shows the response of the 5% longrise.

Figure 11 shows the IRFs of the first- and higher-order moments to a positive shock. For the

standard deviation and tail risk responses, we compute the IRFs as differences in the standard deviation and shortfall and longrise of the $+1$ Shock and Baseline distributions.[9] The standard deviation of the distribution increases, peaking in the first period after the shock. These effects lead to a large increase in the 5% longrise, as a rise in the mean and increase in the standard deviation greatly increases the upside tail risk. The 5% shortfall also increases but by much less, as the mean increase is counteracted by the increased standard deviation. Therefore, through movements in the higher-order moments of the distribution, the model can generate distinct asymmetries in the movements of the upper and lower tails of the distribution, in line with the stylized facts documented by Adrian et al. (2019). These IRFs also reinforce that our model can generate time-varying volatility through the nonlinear dynamics.

Figures 12 and 13 show the corresponding distributional responses to a one standard deviation negative shock. A negative shock lowers the mean and decreases the standard deviation of the distribution. These factors together again generate a more persistent negative movement in the longrise but less persistent effects on the shortfall. Comparing the positive and negative responses, we see that the asymmetries we document for the first moment also carry over to higher moments. The declines in the standard deviation and tail risk are smaller in magnitude when compared with the increases in those features of the distribution following a positive shock.

[9] We define the shortfall of a variable to be $E(x|x < q_\alpha(x))$. In other words, the expectation conditional on that variable being in the bottom α percentile of its distribution. The longrise is $E(x|x > q_{1-\alpha}(x))$, or the expectation conditional on the variable being in the top α percentile of its distribution.

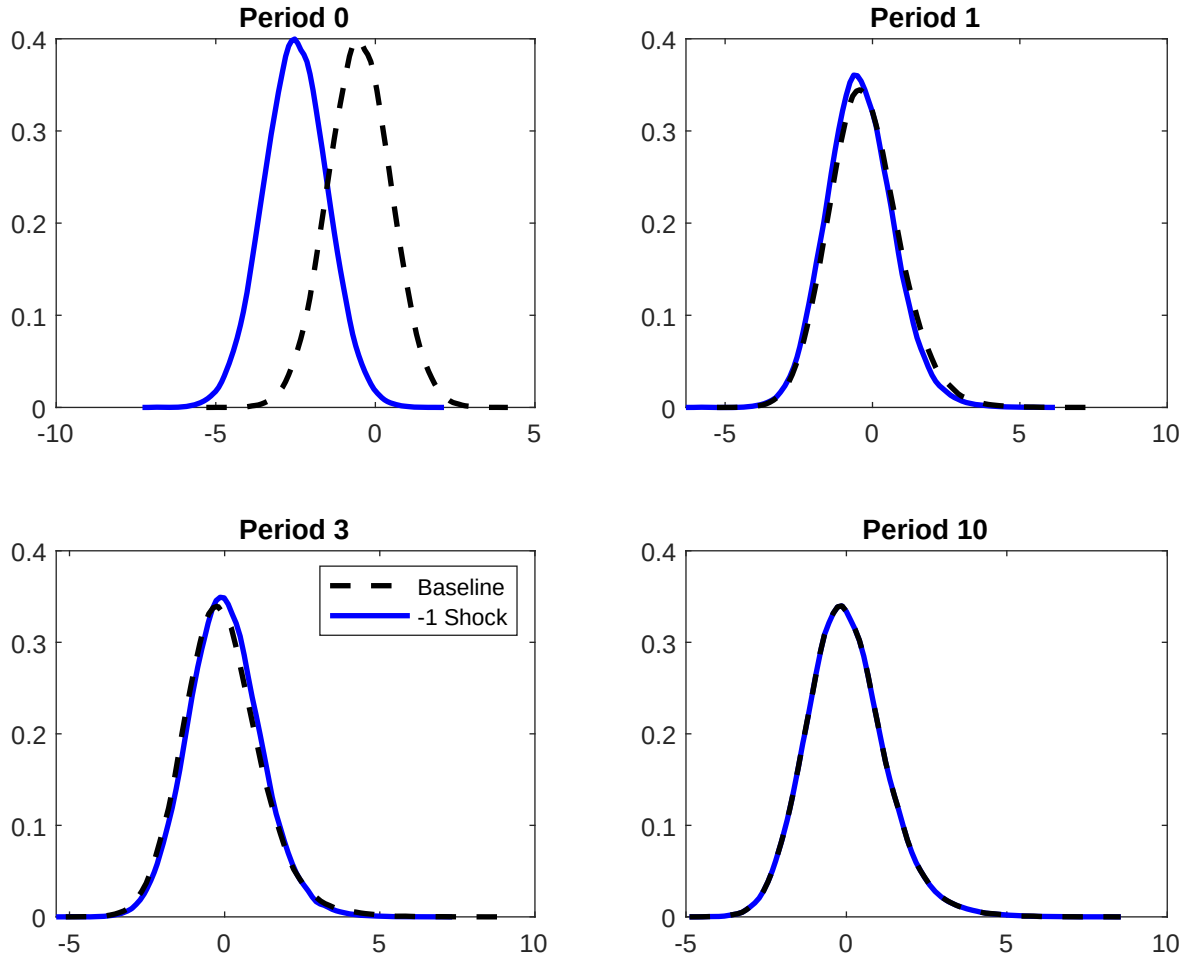


Figure 12: Dynamics of the distribution of the demeaned overall factor at various periods after a shock in a calibrated model. The blue line is the distribution after a negative shock, and the dashed black line is the distribution without a shock. Period 0 is the period of the shock.

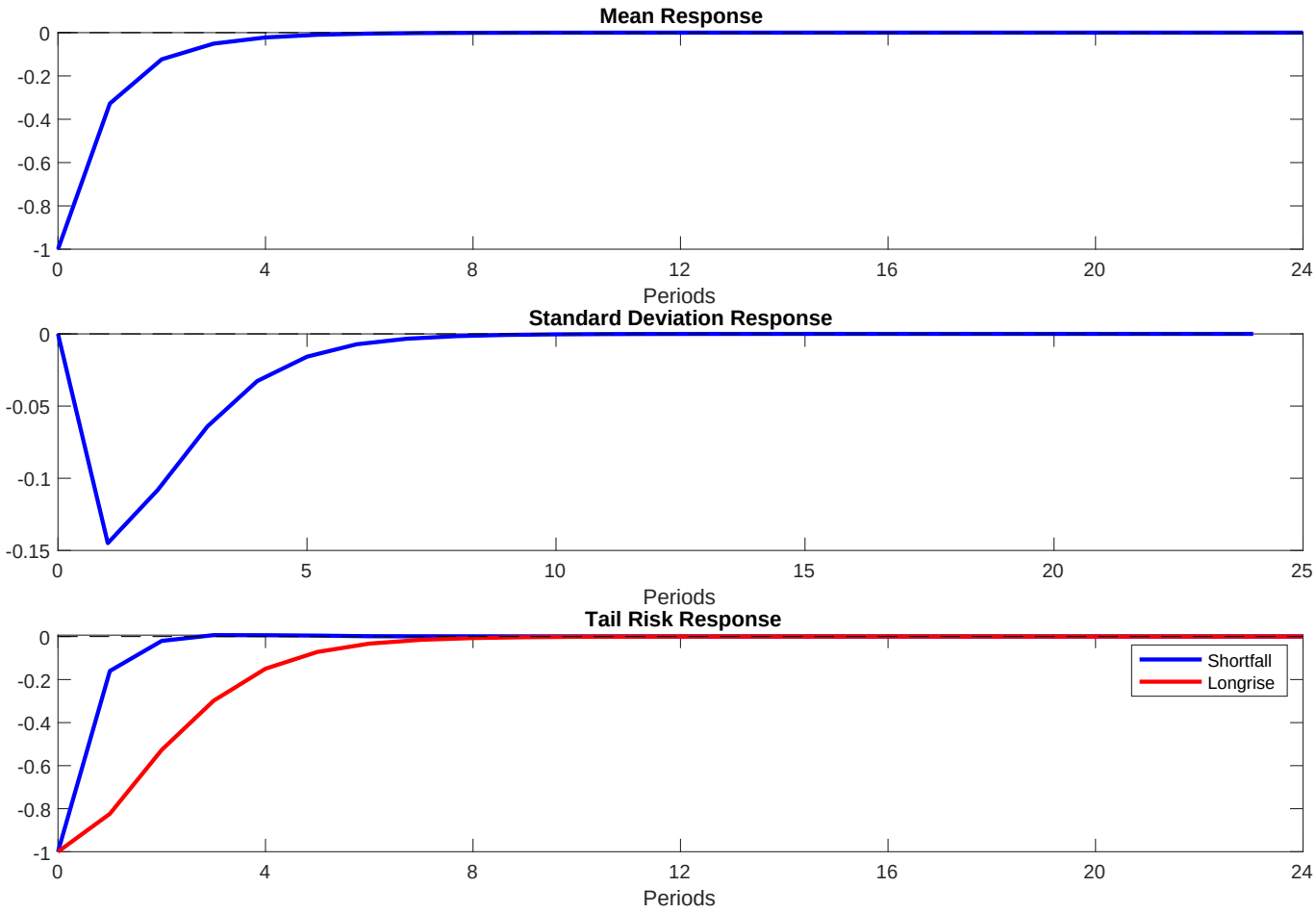


Figure 13: Impulse response functions of the mean, standard deviation, and tail risk of the demeaned overall factor in response to a negative shock at period 0 in a calibrated model. In the third panel, the blue line shows the response of the shortfall, and the red line shows the response of the longrise.

E Estimation Algorithms

In this section of the appendix, we provide details about two estimation algorithms that we use. The first one is a particle Gibbs sampling algorithm. This is the filtering technique used in the main paper. The second one is a Metropolis Hastings algorithm using the particle filter. This approach is used in the application in Appendix Section I. We refer to our benchmark NLDF model, shown again in Equation 21 for convenience.

$$
\begin{aligned}
&y_t = G f_t + \eta \epsilon_t \\
&\begin{cases}
f_t = c + f_t^f + f_t^s \\
f_t^f = h_x f_{t-1}^f + \sigma \nu_t \\
f_t^s = h_x f_{t-1}^s + \frac{1}{2} h_{xx} \left(f_{t-1}^f \right)^2 .
\end{cases}
\end{aligned} \tag{21}
$$

Here, we assume that $c = -\frac{1}{2} \frac{h_{xx}\sigma^2}{(1-h_x)(1-h_x^2)}$.

E.1 Gibbs Sampling with Particle Smoother

Our Gibbs sampling algorithm is as follows:

1. Draw G, η given f_t^f, f_t^s, and y_t. This step follows a standard linear regression model.

2. Draw h_x, h_{xx} given σ, G, η, f_t^f, f_t^s, and y_t.

   We use a random walk Metropolis step to draw h_x and h_{xx}. Given the current accepted draw of h_x and h_{xx}, our proposal is as follows:

$$
\begin{pmatrix} h_x^{prop} \\ h_{xx}^{prop} \end{pmatrix} = \begin{pmatrix} h_x \\ h_{xx} \end{pmatrix} + S^h \zeta, \ \zeta \sim N(0, I).
$$

   We throw away draws that violate the stationarity conditions $h_x^{prop} > 1$.

Given proposed h_x^{prop} and h_{xx}^{prop}, we calculate its likelihood. The new parameters change c and f_t^s.

We update

$$c^{prop} = -\frac{1}{2}\frac{h_{xx}^{prop}\sigma^2}{(1-h_x^{prop})\left(1-(h_x^{prop})^2\right)}$$

and

$$f_t^{s,prop} = h_x^{prop} f_{t-1}^{s,prop} + \frac{1}{2}h_{xx}^{prop}\left(f_{t-1}^f\right)^2.$$

We initialize $f_0^{s,prop} = f_0^s$.

We then form the likelihood of the proposal, which can be calculated in two parts. The first is based on the measurement equation and the second is from the transition equation of the first-order factor:

$$\begin{aligned} y_t - G\left(c^{prop} + f_t^f + f_t^{s,prop}\right) &= \eta\epsilon_t \\ f_t^f - h_x^{prop} f_{t-1}^f &= \sigma\nu_t. \end{aligned} \tag{22}$$

We accept the proposal with probability:

$$prob = \max\left\{\frac{\prod_{t=1}^T p\left(y_t|c^{prop},G,\eta,f_t^f,f_t^{s,prop}\right)p_{trans}\left(f_t^f|h_x^{prop},\sigma,f_{t-1}^f\right)g\left(h_x^{prop},h_{xx}^{prop}\right)}{\prod_{t=1}^T p\left(y_t|c^{curr},G,\eta,f_t^f,f_t^s\right)p_{trans}\left(f_t^f|h_x^{curr},\sigma,f_{t-1}^f\right)g\left(h_x^{curr},h_{xx}^{curr}\right)},1\right\} \tag{23}$$

where $p(y_t|.)$ denotes the likelihood from the measurement equation, $p_{trans}(f_t^f|.)$ denotes the likelihood from the transition equation, and $g(.)$ is the prior distribution.

3. Draw σ^2 given G, η, f_t^f, f_t^s, and y_t.

We draw σ^2 using a random walk Metropolis step. Given the current accepted draw of σ^2, our proposal is as follows:

$$\sigma^{2,prop} = \sigma^2 + S^\sigma \iota, \ \iota \sim N(0, I).$$

We throw out draws that are negative.

Given the proposed $\sigma^{2,prop}$, we calculate its likelihood. The new parameters change c.

We update

$$c^{prop} = -\frac{1}{2} \frac{h_{xx} \sigma^{2,prop}}{(1 - h_x)(1 - h_x^2)}. \tag{24}$$

We then form the likelihood of the proposal, which can be calculated in two parts. The first is based on the measurement equation, and the second is from the transition equation of the first-order factor:

$$\begin{aligned} y_t - G\left(c^{prop} + f_t^f + f_t^s\right) &= \eta \epsilon_t \\ f_t^f - h_x f_{t-1}^f &= \sigma^{prop} \nu_t. \end{aligned} \tag{25}$$

We accept the proposal with probability:

$$prob = \max\left\{ \frac{\prod_{t=1}^T p\left(y_t | c^{prop}, G, \eta, f_t^f, f_t^s\right) p_{trans}\left(f_t^f | h_x, \sigma^{2,prop}, f_{t-1}^f\right) g\left(\sigma^{2,prop}\right)}{\prod_{t=1}^T p\left(y_t | c^{curr}, G, \eta, f_t^f, f_t^s\right) p_{trans}\left(f_t^f | h_x, \sigma^{2,curr}, f_{t-1}^f\right) g\left(\sigma^{2,curr}\right)}, 1 \right\} \tag{26}$$

where $p(y_t|.)$ denotes the likelihood from the measurement equation, $p_{trans}(f_t^f|.)$ denotes the likelihood from the transition equation, and $g(.)$ is the prior distribution.

4. Draw f_t^f, f_t^s given σ, G, η, h_x, h_{xx}, and y_t using the particle Gibbs sampler with ancestor sampling. We discuss our implementation of the sampler here, but further details of the algorithm can be found in Lindsten et al. (2014).

- **Initialize particle smoother:** For particles $j = 1, ..., N-1$. To take a draw from the unconditional distribution, we simulate the model for 500 periods and use the final period of the simulation to determine: $f_0^{f,(j)}, f_0^{s,(j)}, f_1^{s,(j)}$. Note that $f_1^{s,(j)}$ is a function of $f_0^{f,(j)}, f_0^{s,(j)}$, so it is known.

- **Draw first period:** For particles $j = 1, ..., N-1$. We determine $f_1^{f,(j)}, f_2^{s,(j)}$ by simulation.

- **Fix final particle:** Fix $f_0^{f,(N)}, f_0^{s,(N)}, f_1^{f,(N)}, f_1^{s,(N)}$, and $f_2^{s,(N)}$ equal to $f_0^{f,*}, f_0^{s,*}, f_1^{f,*}, f_1^{s,*}$, and $f_2^{s,*}$, where $*$ denotes the accepted previous draw.

- **Set weights:** Compute $w_1^{(j)} = \frac{p(y_1|f_1^{f,(j)}, f_1^{s,(j)})}{\sum_{jj=1}^{N} p(y_1|f_1^{f,(jj)}, f_1^{s,(jj)})}$ for $j = 1, ..., N$.

  For $t = 2, ..., T$:

- **Sample indices to set ancestors for each particle:** For particles $j = 1, ..., N-1$. Draw $a_t^{(j)}$ from the distribution w_{t-1}. Simulate the following:

$$\begin{aligned} f_t^{f,(j)} &= h_x f_{t-1}^{f,(a_t^{(j)})} + \sigma \nu_t \\ f_{t+1}^{s,(j)} &= h_x f_t^{s,a_t^{(j)}} + \frac{1}{2} h_{xx} \left(f_t^{f,(j)} \right)^2 \end{aligned} \tag{27}$$

- **Fix the final particle:** Fix $f_t^{f,(N)}$ equal to $f_t^{f,*}$.

- **Compute auxiliary weights for the fixed particle:** For $j = 1, ..., N$. We compute the auxiliary weights for the fixed particle as follows:

$$w_t^{aux,(j)} = w_{t-1}^{(j)} p(y_t|f_t^{f,(N)}, f_t^{s,(j)}) g(f_t^{f,(N)}|f_{t-1}^{f,(j)}) p(y_{t+1}|f_{t+1}^{f,(N)}, f_{t+1}^{s,(N')}) g(f_{t+1}^{f,(N)}|f_t^{f,(N)}). \tag{28}$$

When calculating $f_{t+1}^{s,(N')}$, we have to take into account that $f_{t+1}^{s,(N')}$ depends on $f_t^{s,(j)}$. Therefore, $f_{t+1}^{s,(N')}$ does not equal $f_{t+1}^{s,(N)}$. The formula is

$$f_{t+1}^{s,(N')} = h_x f_t^{s,(j)} + \frac{1}{2} h_{xx} \left(f_t^{f,(N)} \right)^2. \tag{29}$$

Note that this formula comes from Equation 23 in Lindsten et al. (2014) with lag $= 2$. Our model is a degenerate state-space model discussed in Section 7.2 of that paper. We can view our model alternatively as a non-Markovian model with one factor f_t^f. See the associated discussion there.

- **Sample the associated ancestor index for particle N:** We sample $a_t^{(N)}$ from the distribution w_t^{aux}. Note that we have to update $f_{t+1}^{s,(N)}$ to make it consistent with the selected ancestor:

$$f_{t+1}^{s,(N)} = h_x f_t^{s,(a_t^{(N)})} + \frac{1}{2} h_{xx} \left(f_t^{f,(N)} \right)^2. \tag{30}$$

- **Set weights:** Compute $w_t^{(j)} = \frac{p(y_t | f_t^{f,(j)}, f_t^{s,(j)})}{\sum_{jj=1}^{N} p(y_t | f_t^{f,(jj)}, f_t^{s,(jj)})}$ for $j = 1, ..., N$.
Note that for the $t = T$, we do not have to update f_{t+1}^s because it is the end of the sample. When computing the auxiliary weights for the fixed particle, we also do not consider the $T + 1$ likelihood.

- **Sample selected states:** Sample $*$ according to w_T. Set $f_t^{f,*}, f_t^{s,*}$ equal to the sampled state.

E.2 Metropolis Hastings with Bootstrap Particle Filter

Our Metropolis Hastings algorithm is as follows:

1. Propose a new set of parameters $\Theta^{prop} = \left\{G^{prop}, \eta^{prop}, h_x^{prop}, h_{xx}^{prop}, \sigma^{2,prop}\right\}$.

   - In practice, we break up the proposals into three blocks: Block 1 (factor equation) $\Theta_1^{prop} = \left\{h_x^{prop}, h_{xx}^{prop}, \sigma^{2,prop}\right\}$; Block 2 (measurement equation loadings) $\Theta_2^{prop} = \{G^{prop}\}$; and Block 3 (measurement equation variances) $\Theta_3^{prop} = \{\eta^{prop}\}$. For each block, we take 50 draws, holding the parameters in the other blocks at their previously accepted values. This step is to regularize the sampler and helps with convergence.

$$\Theta_i^{prop} = \Theta_i^{curr} + 0.95 S_{i,1}\zeta_1 + 0.05 S_{i,2}\zeta_2, \ \zeta_i \sim N(0, I) \quad i = 1, 2, 3, 4$$

   - We tune the variance-covariance matrix of the proposals $S_{i,1}$ and $S_{i,2}$ in an adaptive fashion over the first $30,000$ draws of the algorithm. $S_{i,1}$ is calculated using the variance-covariance matrix from all of the previous draws multiplied by a scaling parameter that decreases if the previous 250 draws within the block had an acceptance rate less than 10%. $S_{i,2}$ is a diagonal matrix that is meant to introduce some independent noise within the proposal. It is multiplied by a separate scaling parameter that decreases if the previous 250 draws within the block had an acceptance rate less than 10%.

2. Evaluate the likelihood of the proposed parameters using the bootstrap particle filter (Särkkä, 2013).

   - **Initialize the particle filter:** For particles $j = 1, ..., N$. To take a draw from the unconditional distribution, we simulate the model for 500 periods and use the final period of the simulation to determine: $f_0^{f,(j)}, f_0^{s,(j)}, f_1^{s,(j)}$. Note that $f_1^{s,(j)}$ is a function of $f_0^{f,(j)}, f_0^{s,(j)}$, so it is known. We set $w_t^{(j)} = 1$ for all particles.

     For $t = 1, ..., T$:

   - **Prediction step:**

     Given particles and weights at $t-1$: $\left\{f_{t-1}^{f,(j)}, f_t^{s,(j)}, w_{t-1}^{(j)}\right\}$.

(a) For particles $j = 1, ..., N$. Draw a new particle $\left\{ f_t^{f,(j)}, f_{t+1}^{s,(j)} \right\}$ from

$$f_t^{f,(j)} = h_x f_{t-1}^{f,(j)} + \sigma \nu_t$$
$$f_{t+1}^{s,(j)} = h_x f_t^{s,(j)} + \frac{1}{2} h_{xx} \left(f_t^{f,(j)} \right)^2 .$$

(b) Calculate weights:

$$\omega_t^{(j)} = p(y_t | f_t^{f,(j)}, f_t^{s,(j)}), \quad j = 1, \ldots, N.$$

- **Update step:**

  (a) Define normalized weights: $\widetilde{w}_t^{(j)} = \frac{\omega_t^{(j)} w_{t-1}^{(j)}}{\frac{1}{N} \sum \omega_t^{(j)} w_{t-1}^{(j)}}$.

  (b) Resample from multinomial distribution $\left\{ \omega_t^{(j)}, \widetilde{w}_t^{(j)} \right\}$ and set $w_t^{(j)} = 1$.

- **Compute conditional likelihood:**

$$p(y_t | Y_{1:t-1}) \approx \frac{1}{N} \sum_{i=1}^{N} \omega_t^{(j)} w_{t-1}^{(j)}. \tag{31}$$

The overall likelihood is then $p(y | \Theta_i^{prop}, \Theta_{-i}^{curr}) = \prod_{t=1}^{T} p(y_t | Y_{1:t-1})$.

3. We accept the proposal with probability

$$prob = \max \left\{ \frac{p(y | \Theta_i^{prop}, \Theta_{-i}^{curr}) g(\Theta_i^{prop}, \Theta_{-i}^{curr})}{p(y | \Theta_i^{curr}, \Theta_{-i}^{curr}) g(\Theta_i^{curr}, \Theta_{-i}^{curr})}, 1 \right\} \tag{32}$$

where $g(.)$ is the prior distribution.

F Monte Carlo Results

We discuss the results from three Monte Carlo experiments. First, we compare a correctly-specified NLDF model with a misspecified linear model. Second, we verify our identification results in a two factor model. Finally, we do the same for the model with separate weights on the first- and second-order factors.

F.1 Correctly-Specified NLDF Model vs. Misspecified Linear Model

We show that if the true data-generating process is the NLDF model, our estimation strategy successfully recovers all parameters. We assume that the underlying model is our benchmark NLDF model with the following parameters: $c = 0, h_x = 0.85$, $h_{xx} = 2.15$, $\sigma = 0.18$, $\text{diag}(\eta) = [0.54, 0.06, 0.79, 1.08, 0.39]$, and $G = [1, 0.17, 1.5, 2.21, 0.56]$. We generate 50 series of length $T = 1000$, starting from $f_0^f = 0, f_0^s = 0$.

With the synthetic data in hand, we then estimate the linear factor model and our benchmark NLDF model with a linear measurement equation. The models are estimated using the Metropolis Hastings and particle filter procedure detailed in Section 2.5 with $200,000$ Markov Chain Monte Carlo draws. We assume flat priors for all of the parameters.

The parameter estimates converge to the true values under correct specification. As seen in the left panel in Figure 14, the log likelihood is higher for the nonlinear model (vertical axis) than it is for the linear one (horizontal axis) across all simulations. The average difference between the log likelihoods in the nonlinear and linear models is 80 points; the difference can be as low as 48 points and as high as 127 points. Correspondingly, the mean square errors of the factors are smaller in the nonlinear factor version (right panel in Figure 14).[10]

We report the estimates of the state equation's parameters in Figures 15 and 16. Whereas the nonlinear factor model's estimate for h_x (y-axis) is clustered around its true value, the linear estimate (x-axis) is about 14% more persistent. This over-persistence is compensated for with a downward bias estimate of the factor innovation volatility. This compensation is needed so the factor

[10]The mean square error is defined as $\frac{\sum_{t=1}^{T=1000}(\hat{f}_{t|t}-f_t)^2}{T}$, where $\hat{f}_{t|t}$ is the factor filtered from the estimated model (linear or nonlinear), and f_t is the true simulated factor.

delivers second moments consistent with the data. In contrast, the volatility estimate from the NLDF model is around the true value. Furthermore, the second-order component (h_{xx}) is estimated close to its true value.

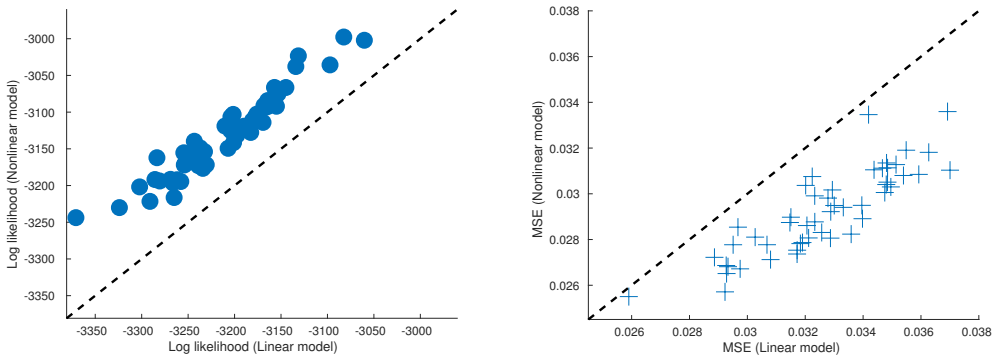


Figure 14: Performance of Estimated Linear and Nonlinear Models on Simulated Data

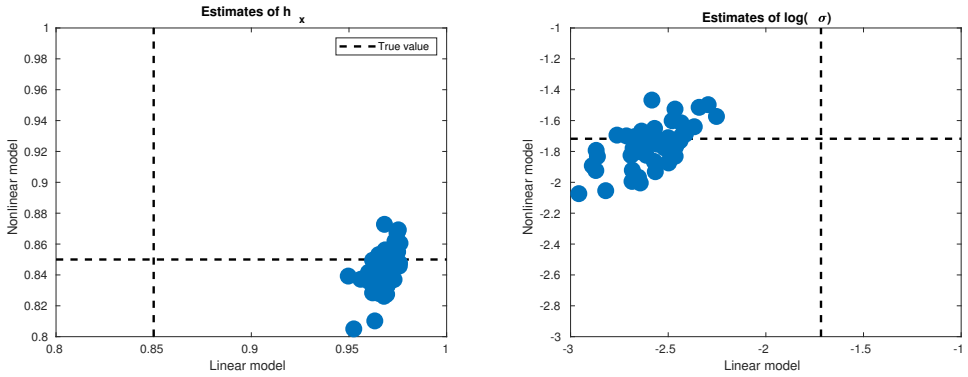


Figure 15: Estimation Bias of Linear and Nonlinear Models on Simulated Data

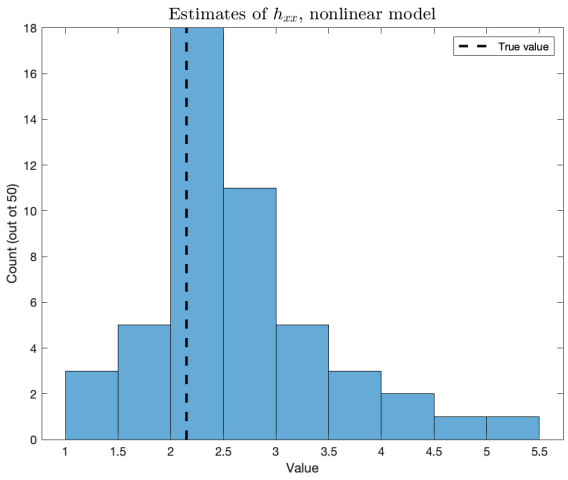


Figure 16: Performance of Nonlinear Model in Estimation of h_{xx} on Simulated Data

F.2 Two Factor Model

We simulate five samples of data of length 500 from a two factor version of our model with the parameterization set in the second column of Table 1.[11]

Our parameterization for the simulation is consistent with our theoretical results in that we are imposing a 2×2 lower triangular structure with ones on the main diagonal in the factor loading matrix combined with a diagonal variance covariance matrix for the shocks to the factors. This is one of the identifications discussed in Bai and Wang (2015) for linear dynamic factor models that also applies in our case, as discussed in Appendix Section A.

We run our Gibbs sampling with particle smoother estimation on the simulated data. Table 1 shows the 95% credible sets for the model parameters. The estimates in bold denote cases where the 95% credible sets do not contain the true parameters. As expected from an identified model, the 95% credible sets contain the true parameters in the majority of cases.

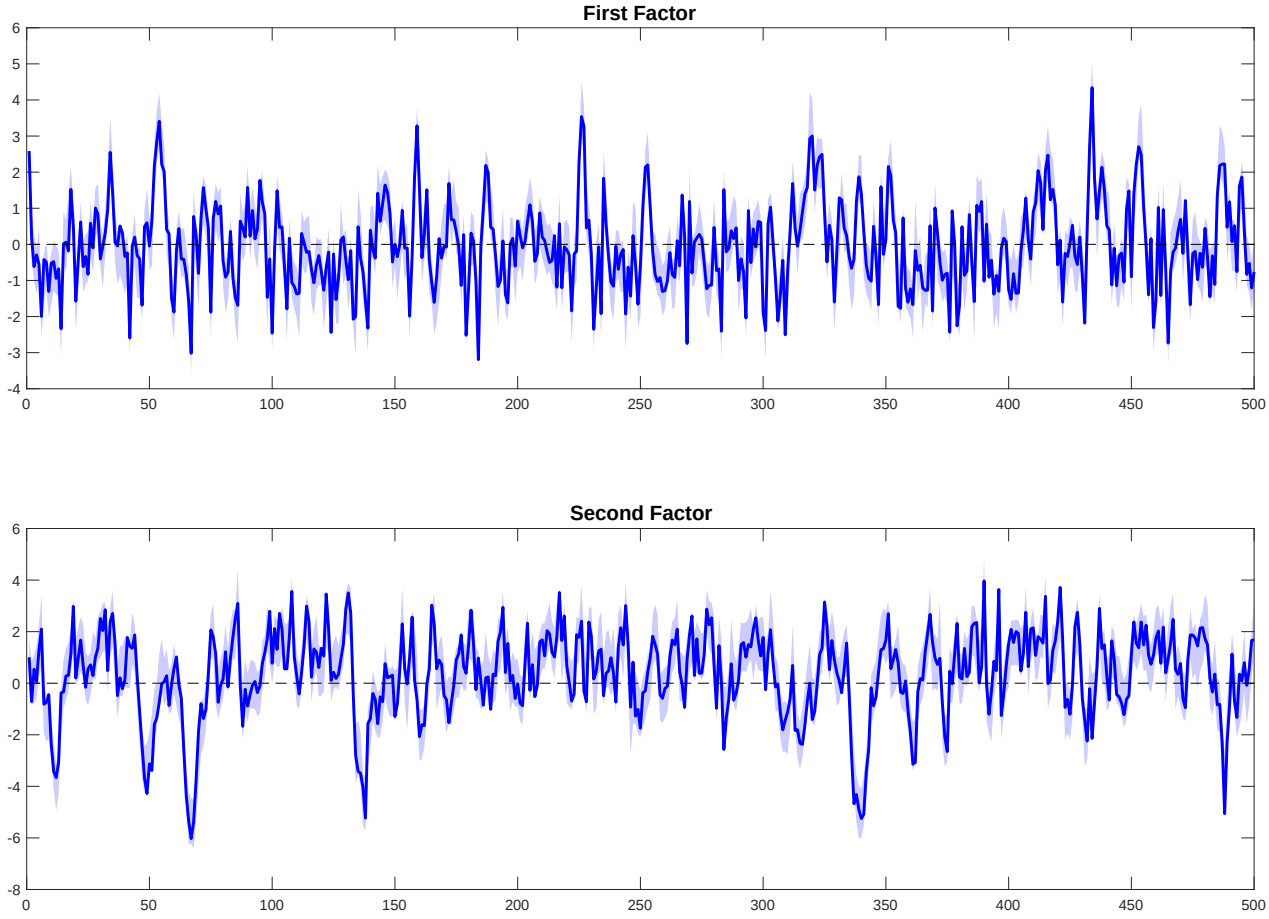
[11] We set the h_x and σ matrices to be diagonal with elements $h_{x,i}$ and σ_i, respectively for $i = 1, 2$. We also specify that $h_{xx,i}$ is univariate and only loads onto $\left(f^f_{i,t-1}\right)^2$. In other words, for both factors, we set the dependence on $f_{1,t-1}f_{2,t-1}$ to be 0. We set $c_1 = -\frac{1}{2}\frac{h_{xx,1}\sigma_1^2}{(1-h_{x,1})(1-h_{x,1}^2)}$ and $c_2 = -\frac{1}{2}\frac{h_{xx,2}\sigma_2^2}{(1-h_{x,2})(1-h_{x,2}^2)}$.

Table 1: Parameter Estimates: 2.5%, Median, and 97.5% Quantiles Across Five Simulations

Parameter		Simulation 1			Simulation 2			Simulation 3			Simulation 4			Simulation 5		
	True	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
$h_{x,1}$	0.45	0.38	0.46	0.52	0.41	0.48	0.56	0.33	0.41	0.47	0.36	0.43	0.50	0.40	0.46	0.52
$h_{xx,1}$	0.5	0.40	0.57	0.79	0.31	0.44	0.58	0.21	0.35	0.50	0.31	0.43	0.55	0.41	0.55	0.68
σ_1^2	1	0.82	0.97	1.16	0.90	1.07	1.27	0.77	0.92	1.09	0.84	1.00	1.17	0.81	0.96	1.13
$h_{x,2}$	0.7	0.63	0.68	0.73	0.67	0.72	0.77	0.67	0.71	0.74	0.63	0.68	0.71	0.64	0.68	0.72
$h_{xx,2}$	-0.5	-0.99	-0.62	-0.38	-0.57	-0.37	-0.21	-0.61	-0.43	-0.29	-0.92	-0.57	-0.38	-0.75	-0.47	-0.28
σ_2^2	1	0.71	0.95	1.22	0.97	1.22	1.46	0.82	1.00	1.21	0.78	1.02	1.33	0.90	1.12	1.40
$G_{1,1}$	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$G_{2,1}$	2	1.73	1.91	2.05	1.86	2.00	2.14	1.92	2.10	2.27	1.84	2.01	2.17	1.82	1.95	2.08
$G_{3,1}$	0.5	0.36	0.43	0.52	0.39	0.46	0.53	0.40	0.49	0.58	0.40	0.48	0.55	0.42	0.49	0.57
$G_{4,1}$	0.6	0.44	0.54	0.65	0.51	0.62	0.74	0.48	0.60	0.73	0.44	0.56	0.68	0.49	0.60	0.73
$G_{1,2}$	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$G_{2,2}$	-0.5	-0.61	-0.52	-0.43	-0.61	-0.54	-0.47	-0.61	-0.54	-0.46	-0.60	-0.51	-0.43	-0.65	-0.55	-0.46
$G_{3,2}$	0.7	0.66	0.71	0.77	0.64	0.68	0.72	0.69	0.73	0.77	0.67	0.71	0.75	0.67	0.72	0.77
$G_{4,2}$	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
η_1^2	0.5	0.40	0.49	0.60	0.40	0.49	0.58	0.39	0.48	0.56	0.43	0.52	0.62	0.40	0.48	0.56
η_2^2	0.75	0.56	0.90	1.33	0.34	0.64	1.02	0.23	0.51	0.87	0.33	0.65	1.00	0.57	0.87	1.22
η_3^2	0.3	**0.32**	**0.40**	**0.50**	0.28	0.35	0.43	0.22	0.29	0.36	0.20	0.26	0.33	0.19	0.26	0.34
η_4^2	0.9	0.60	0.76	0.94	**0.55**	**0.70**	**0.86**	0.63	0.79	0.95	0.69	0.86	1.02	0.78	0.92	1.08

Figure 17 shows a comparison of the true factors in solid blue against the 95% credible bands of the smoothed estimates from Simulation 1. The smoothed estimates mostly contain the true values, which is further evidence that the parameter restrictions are sufficient to identify the factors.

Figure 17: Estimated Factors Versus True Factors: Two Factor Model



NOTE: This estimation is from Simulation 1. The blue solid lines are the true factors. Shaded areas are the 95% credible sets of the smoothed estimates of the two factors.

F.3 Separate Loadings on First- and Second-Order Components

We simulate five samples of data of length 500 from a version of our model with different loadings on the first- and second-order components. We have one first-order and one second-order component. The parameterization for the simulation is shown in the second column of Table 2.

Our parameterization specifies a unit factor loading on the first observable for both the first- and second-order components. This is consistent with the identifying restrictions discussed in Appendix Section A.

As Table 2 shows, across the five simulations, the true parameters are always contained within the 95% credible sets.

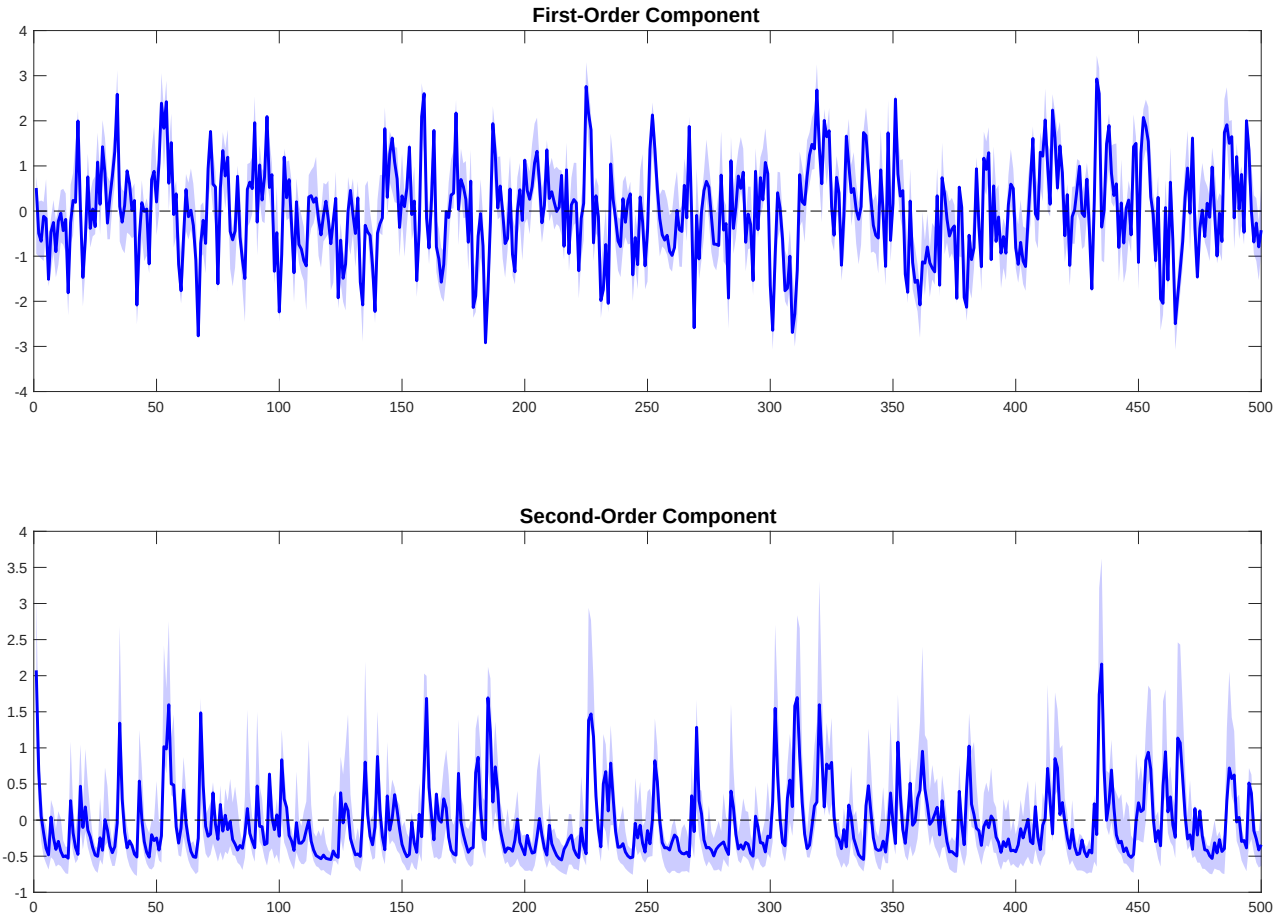
Table 2: Parameter Estimates: 2.5%, Median, and 97.5% Quantiles Across Five Simulations

Parameter		Simulation 1			Simulation 2			Simulation 3			Simulation 4			Simulation 5		
	True	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
h_x	0.45	0.40	0.46	0.51	0.39	0.45	0.50	0.34	0.41	0.47	0.36	0.43	0.49	0.40	0.45	0.50
h_{xx}	0.5	0.50	0.64	0.80	0.40	0.50	0.60	0.40	0.53	0.66	0.32	0.44	0.56	0.43	0.54	0.68
σ^2	1	0.79	0.93	1.11	0.87	1.01	1.17	0.84	0.97	1.12	0.88	1.04	1.20	0.81	0.95	1.11
$G_{1,1}$	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$G_{2,1}$	2	1.79	1.96	2.17	1.81	1.95	2.11	1.85	2.00	2.18	1.78	1.94	2.11	1.84	1.99	2.17
$G_{3,1}$	0.50	0.44	0.51	0.56	0.41	0.46	0.51	0.43	0.49	0.55	0.43	0.48	0.54	0.43	0.48	0.53
$G_{4,1}$	0.60	0.51	0.60	0.69	0.55	0.63	0.70	0.55	0.65	0.74	0.46	0.56	0.64	0.56	0.66	0.74
$G_{1,2}$	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$G_{2,2}$	-0.5	-0.63	-0.30	-0.05	-0.72	-0.44	-0.20	-0.84	-0.38	-0.05	-1.01	-0.55	-0.22	-0.66	-0.39	-0.18
$G_{3,2}$	0.7	0.47	0.58	0.70	0.54	0.66	0.78	0.55	0.68	0.84	0.54	0.69	0.88	0.61	0.71	0.82
$G_{4,2}$	0.9	0.65	0.81	1.01	0.74	0.90	1.10	0.53	0.74	0.96	0.73	0.95	1.23	0.66	0.82	0.97
η_1^2	0.5	0.42	0.50	0.60	0.38	0.46	0.55	0.35	0.43	0.51	0.43	0.52	0.62	0.41	0.49	0.57
η_2^2	0.75	0.69	0.88	1.11	0.62	0.82	1.05	0.66	0.88	1.09	0.68	0.89	1.14	0.65	0.84	1.08
η_3^2	0.3	0.27	0.31	0.37	0.27	0.31	0.36	0.25	0.29	0.33	0.26	0.30	0.35	0.24	0.28	0.33
η_4^2	0.9	0.72	0.81	0.94	0.73	0.82	0.94	0.73	0.82	0.95	0.79	0.90	1.04	0.76	0.86	1.00

We set $h_{x,1} = h_{x,2} = h_x$.

Figure 18 shows the true first- and second-order components in solid blue and the 95% credible set estimates in shaded blue for Simulation 1. In both cases, the credible sets often cover the true values.

Figure 18: Estimated Factor Versus True Factor: First- and Second-Order Components



NOTE: This estimation is from Simulation 1. The blue solid lines are the true first- and second-order components. Shaded areas are the 95% credible sets of the smoothed estimates of the first- and second-order components.

G Model with Quadratic Measurement Equation

We compare our baseline NLDF model to one with a quadratic measurement equation and linear factor dynamics. We show a univariate version of the model in Equation 33:

$$\begin{aligned} y_t &= G_1 f_t + G_2 \left(c + f_t^2\right) + \eta \epsilon_t \\ f_t &= h_x f_{t-1} + \sigma \nu_t. \end{aligned} \tag{33}$$

We set $c = -\frac{\sigma^2}{1-h_x^2}$ to ensure that the unconditional mean of y_t is 0.

The factor f_t and squared factor f_t^2 can be written in a state space form shown in Equation 34. Compared to Equation 10 for the NLDF model, this state space does not contain the terms for the second-order factor. However, the rest of the elements in the state vector are the same. Another key difference is that the f_t^2 term impacts the observables y_t directly and with a different factor loading compared to f_t.

$$\begin{pmatrix} f_t \\ f_t^2 \end{pmatrix} = \begin{pmatrix} h_x & 0 \\ 0 & h_x^2 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-1}^2 \end{pmatrix} + \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma^2 & 2\sigma h_x \end{pmatrix} \begin{pmatrix} \nu_t \\ \nu_t^2 \\ f_{t-1}\nu_t \end{pmatrix} \tag{34}$$

Mirroring our discussion of the NLDF model, when we discuss conditional moments at time t, the conditioning set is the past history of factors $\{f_t, f_{t-1}, ...\}$.

G.1 First Moment Dynamics

Equation 35 shows the $h-$step ahead conditional expectations dynamics of y_t. The first term in the final expression is determined by the factor f_t and it decays in the horizon h as in a linear dynamic factor model. The second expression depends on the squared factor f_t^2. As the horizon h increases, there are two counterbalancing effects. First, the term $\sigma^2 \frac{1-h_x^{2h}}{1-h_x^2}$ increases, which comes from the expectation of additional squared innovations $\nu_{t+1}^2, ..., \nu_{t+h}^2$. As the horizon tends to ∞, this term converges to $-c$. Second, the last term in the parenthesis that depends on f_t^2 decays

according to h_x^{2h}.

$$E_t\left(y_{t+h}\right) = G_1 E_t\left(f_{t+h}\right) + G_2 E_t\left(c + f_{t+h}^2\right) = G_1 h_x^h f_t + G_2\left(c + \sigma^2 \frac{1 - h_x^{2h}}{1 - h_x^2} + h_x^{2h} f_t^2\right) \tag{35}$$

G.2 Impulse Response Functions

The impulse response functions in this model are asymmetric, state-dependent, and size-dependent for the same reasons as in the NLDF model. Looking at the state space in Equation 34, this is because the innovations to f_t^2 are of the same form.

In this model, we can more easily write down a closed-form formula for the $h-$step ahead impulse response function of y_t with respect to a shock of size δ at time t and given time $t-1$ conditions. This is shown in Equation 36. The first term in the final expression is the same as in a linear dynamic factor model. This part of the IRF leads to symmetric, state-independent, and size-independent responses to shocks.

$$\begin{aligned} IRF_{\delta,t-1}\left(y_{t+h}\right) &= E\left(y_{t+h}(\delta) - y_{t+h}|\Omega_{t-1}\right) \\ &= G_1 h_x^h \sigma\delta + G_2 h_x^{2h} E_{t-1}\left[\left(\left(h_x f_{t-1} + \sigma\left(\nu_t + \delta\right)\right)^2 - \left(h_x f_{t-1} + \sigma\nu_t\right)^2\right)\right] \\ &= G_1 h_x^h \sigma\delta + G_2 h_x^{2h}\left(2h_x f_{t-1}\sigma\delta + \sigma^2\delta^2\right) \end{aligned} \tag{36}$$

The second term provides an additional perspective on how the model generates the nonlinearities in the IRF. First, the δ^2 term induces an asymmetric response to positive versus negative shocks. While positive and negative shocks have effects of equal magnitudes and opposite signs in the first two terms of the IRF, thereby leading to symmetry in the IRF from these terms, the third one always increases no matter the direction of the shock. Second, the presence of f_{t-1} in the IRF generates state dependence, as the sign and magnitude of the lagged factor influence the shape of the IRF. Finally, there is size dependence through the δ^2 term, as larger shocks have larger-than-proportionate effects on the observables compared to smaller ones. The overall impact on observables is modulated through G_2, which is the factor loading on the squared factor term.

Moreover, the dependence of the nonlinear terms on the persistence parameter h_x also means that these effects are monotonically decreasing with the horizon h.

G.3 Volatility Dynamics

The model with the quadratic measurement equation generates time-varying volatility that depends on the level of the factor f_t. This can be seen from Equation 34 by noting that the impact of the innovation ν_t on f_t^2 depends on f_{t-1}. This is the same mechanism as discussed in the NLDF model, with the difference that f_t^2 directly enters the measurement equation instead of its lags feeding through the second-order factor f_t^s.

$$\begin{aligned} Var_t\left(y_{t+h}\right) &= G_1^2 Var_t\left(f_{t+h}\right) + G_2^2 Var_t\left(f_{t+h}^2\right) + 2G_1G_2 Cov_t\left(f_{t+h}, f_{t+h}^2\right) + \eta^2 \qquad (37) \\ &= G_1^2\sigma^2\frac{1-h_x^{2h}}{1-h_x^2} + 2G_2^2\sigma^2\left[\sigma^2\left(\frac{1-h_x^{2h}}{1-h_x^2}\right)^2 + 2h_x^{2h}\frac{1-h_x^{2h}}{1-h_x^2}f_t^2\right] + 4G_1G_2\sigma^2 h_x^h\frac{1-h_x^{2h}}{1-h_x^2}f_t + \eta^2 \end{aligned}$$

The simpler structure of the model with the quadratic measurement equation allows us to compute a formula for the $h-$step ahead conditional variance of observables, which is shown in Equation 37. This formula is broken up into four components. The first one depends on the conditional variance of f_{t+h} and is the same as in a standard linear dynamic factor model. It monotonically increases with the horizon h and does not show time-varying volatility. The last term is from the measurement error and has constant variance. The time-varying volatility terms come from the second and third terms, which depend on f_t^2 and f_t, respectively.

We begin by discussing the second term, which depends on f_t^2 and has a more straightforward interpretation. This term captures the conditional variance of f_{t+h}^2 and depends on f_t^2. If f_{t+h}^2 is larger, which occurs when f_{t+h} is further away from 0, then the effects of a change in f_{t+h} on f_{t+h}^2 are amplified and the conditional variance is increased. Because of persistence in f_t given $h_x > 0$, f_{t+h}^2 tends to be large in magnitude when f_t^2 is large in magnitude. These effects depend on the horizon h in potentially nonlinear ways depending on the magnitude of h_x, although they eventually dissipate to 0 as the horizon lengthens.

The level of the factor f_t, which determines the third term in the final expression, can either

increase or decrease volatility depending on its sign and the signs of the factor loadings. This term captures the covariance between f_{t+h} and f_{t+h}^2. If G_1 and G_2 have the same sign, then a positive f_t increases the conditional variance while a negative f_t decreases it. The intuition is as follows. The factor f_{t+h} positively comoves with f_{t+h}^2 when f_{t+h} is positive, which is more likely to occur when f_t is positive due to persistence, assuming $h_x > 0$. If G_1 and G_2 also have the same sign, then the positive comovements in f_t and f_{t+h}^2 have amplifying effects on the observables and the conditional variance increases. If G_1 and G_2 have opposite signs, then the movements of f_{t+h} and f_{t+h}^2 partially cancel each other out and the conditional variance decreases. When f_{t+h} is negative, then its comovement with f_{t+h}^2 is negative and the effects are reversed.

G.4 Relationship Between First and Second Moments

The presence of f_t and f_t^2 in the conditional mean and variance equations indicates that the model can generate comovement between first and second moment dynamics, thereby leading to growth-at-risk features and other non-normalities in the conditional distribution of the observables. As f_t and f_t^2 are positively correlated, the key determinants for the direction of comovement are the signs of the factor loadings G_1 and G_2. If $G_2 > 0$, then an increase in f_t^2 both increases the conditional mean and variance while if $G_2 < 0$, then an increase in f_t^2 decreases the conditional mean while increasing the conditional variance. If G_1 is the same sign as G_2, then f_t works in the same direction as f_t^2 on the conditional variance. If both factor loadings are positive, then there is a positive comovement between mean and variance while if both are negative, then there is a negative comovement. If G_1 and G_2 are of opposite signs, then the two components' movements partially cancel each other out in both the mean and variance.

G.5 Comparison Between NLDF and Quadratic Measurement Equation Models

The NLDF and quadratic measurement equation models share many features, which explains the similarities in many of the first and second moment properties. Most importantly, the heart of the nonlinearity in the models both come from the squared factor. Comparing the models, however,

there are two important differences.

First, the squared factor enters the system in different ways. In the NLDF model, the lagged squared component impacts the second-order component, which itself propagates with persistence h_x. The second-order component affects the overall factor and therefore the observables. By repeated substitution, we see that the second-order component at time t is a function of the squared first-order component at time $t-1$ and further lags:

$$f_t^s = \frac{1}{2} h_{xx} \left(\left(f_{t-1}^f \right)^2 + h_x \left(f_{t-2}^f \right)^2 + h_x^2 \left(f_{t-3}^f \right)^2 + ... \right).$$

In contrast, the quadratic measurement equation model specifies the time t observable to depend on the contemporaneous squared factor.

This means that in the NLDF model, the nonlinear implications of a shock begin to impact the overall factor with a lag of one period. Therefore, the one-step ahead conditional distributions generated by the model are normally distributed. If the measurement equation is linear, as in our baseline specification, then this conditional normality one step ahead propagates through to the observables as well. The persistence of the second-order component also allows its dynamics to meaningfully deviate from those of the first-order component. The quadratic measurement equation does not have this additional source of persistence and so the squared factor is tightly linked to the contemporaneous level factor. However, it generates a non-Gaussian conditional distribution of the observables even one-step ahead.

A second key difference is that in the NLDF model, the second-order component impacts the observables with the same loading as the first-order component while in the quadratic measurement equation model, the quadratic factor impacts the observable with a potentially different weight when compared to the linear factor. In the quadratic measurement equation model, the factor loading on the quadratic factor plays a similar role to the h_{xx} in the NLDF model, the latter which governs the relationship between the second-order component and first-order component squared. In Section 4.4 of the main draft, we explore the implications of allowing for separate factor loadings in the first- and second-order components of the NLDF model.

H Nonlinear Credit Cycle: Additional Results

H.1 Details on Data Construction

Our data extend from 1952:Q1 through 2021:Q4 at a quarterly frequency. Our data are from the Statistical Release Z.1 "Financial Accounts of the United States" and were downloaded from the Federal Reserve Bank of St. Louis, Federal Reserve Economic Data. These data are not seasonally adjusted, and we seasonally adjust them using the Census X-13 Seasonal Adjustment procedure implemented in Eviews 12. We deflate the seasonally adjusted data by the seasonally adjusted GDP deflator to turn them into real values.

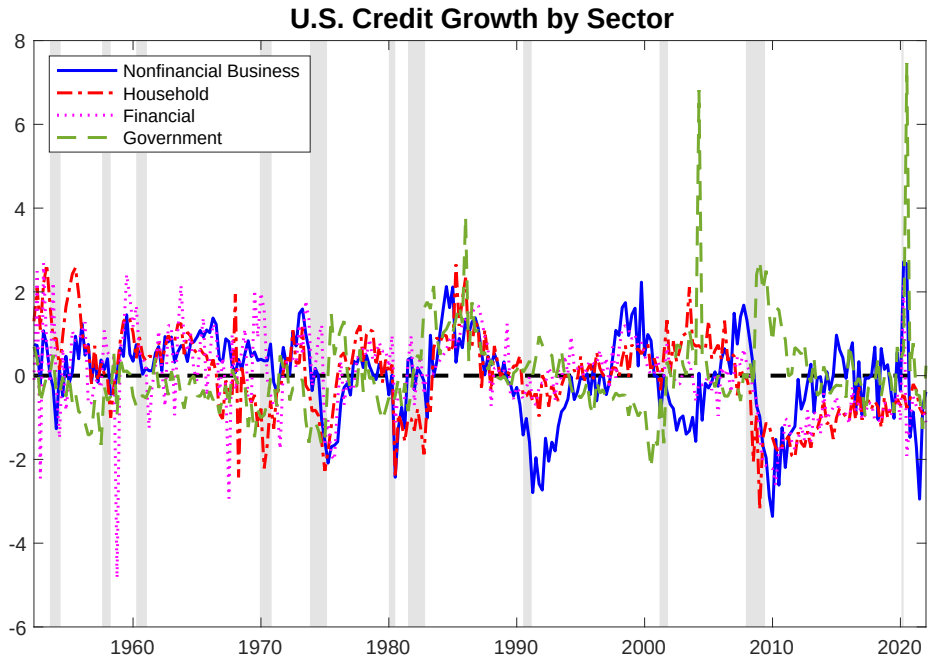


Figure 19: Normalized real credit growth by sector in the United States: 1952:Q1-2021:Q4 with National Bureau of Economic Research recession shading.

For nonfinancial business debt, we use the category Nonfinancial Business, Debt Securities and Loans, Liability, Level (BOGZ1FL144104005Q). For household debt, we use the category Households and Nonprofit Organizations, Debt Securities and Loans, Liability, Level (TCMILBSHNO). For financial-sector debt, we use the category Domestic Financial Sectors, Debt Securities and Loans, Liability, Level (TCMDODFS). Finally, for government debt, we sum the categories Federal Government, Debt Securities and Loans, Liability, Level (FGTCMDODNS) and State and Lo-

cal Governments, Debt Securities and Loans, Liability, Level (SLGTCMDODNS). We seasonally adjust the federal and state and local government debt separately before summing them up.

H.2 Credit Growth Dynamics

Figure 19 shows the data that we use to estimate the model, which are normalized U.S. real credit growth in the nonfinancial business, household, financial, and government sectors. Private credit growth generally increases during expansions and declines during recessions, although the troughs of nonfinancial business and financial credit growth lag the troughs of recessions. The three private credit growth series are fairly positively correlated, ranging from 0.4 to 0.5. In the nonfinancial business and household sectors, credit growth exhibits an important asymmetry, with expansions marked by steady, strong growth and recessions associated with sharp, violent declines. These dynamics have implications for higher-order moments, with the Kelly skewness of nonfinancial business credit growth at -0.23 and household credit growth at -0.15. Financial credit growth experienced rapid declines in the GFC but overall has a skewness close to 0.

Conversely, government credit growth is mildly negatively correlated to the three other series because it has increased in recent recessions. The series has a distinct positive skew due to several large spikes in public debt.

H.3 Unconditional Distribution of the Factors Implied by the Model

A key implication of the nonlinear model is that the unconditional distribution of the factor is not normally distributed, even though the exogenous innovations to the system are. This divergence from normality does not occur if we ignore the second-order component. In examining the credit growth data, we saw some evidence of asymmetries. These features of the data inform the estimation of the nonlinear model. Table 3 shows that the nonlinear model generates a negative Kelly skewness, with mass below the median of the distribution covering nearly 60% of the total distance from the 10th to the 90th percentiles. The credible sets of the skewness estimates are wide, reflecting the difficulty in pinning down the magnitude of the higher-order moments. However, the evidence indicates that the skewness is negative at the 68th percentile credible sets, as seen in

Table 3: Unconditional Moments Implied by the Nonlinear Dynamic Factor Model and the Linear Factor Only

	Skewness	5% Shortfall	95% Longrise	Corr$\left(f^s, \left(f^f\right)^2\right)$	Variance Decomp
Nonlinear	-0.18	-1.91	1.22	-0.53	9.69
	$[-0.27, -0.07]$	$[-2.59, -1.38]$	$[1.08, 1.37]$	$[-0.54, -0.51]$	$[1.41, 25.78]$
Linear Only	-0.00	-1.34	1.33	–	–
	$[-0.00, 0.00]$	$[-1.49, -1.20]$	$[-1.48, -1.19]$	–	–

NOTE: The table shows the Kelly skewness, 5th/95th shortfall and longrise, and the variance decomposition showing the percentage of unconditional variation implied by the second-order factor. "Nonlinear" refers to the full model while "Linear Only" refers to a counterfactual in which $h_{xx} = 0$ for all of the draws, keeping everything else the same. The headline number is the posterior median while the numbers in brackets are the $16/84$ credible sets.

the table. This behavior continues to be the case at the 80th percentile sets as well. The second row of the table shows the corresponding estimates for the linear-only model. With a linear process and Gaussian shocks, the model cannot generate any skewness.

The next two columns in Table 3 show the estimates of the lower and upper tails of the distribution. As a reference, the mean of the factor by assumption is 0. The nonlinear model generates a distribution that has higher probability on large declines in the credit cycle as opposed to large increases. This asymmetric tail behavior is consistent with the negative skewness previously discussed. In comparison, the linear model generates symmetric tail behavior.

Underlying the skewness and tail risk behavior of the model is a strong correlation between the level and volatility components of the nonlinear factor. The second-to-last column of Table 3 shows the model-implied correlation between the second-order factor, which enters into the level of the nonlinear factor, and the square of the first-order factor, which determines the conditional volatility of the innovations to the second-order factor. This correlation is -0.53, which suggests that the conditional volatility of the credit cycle increases as the credit cycle declines. This result is again consistent with the idea that credit expansions are smoother than credit contractions. Moreover, it can also generate the negative skewness and long lower tails coupled with short upper tails we see.

In addition to examining deviations from normality, we can compute the unconditional variance decomposition of the overall factor into its linear and nonlinear components. If the second-

order factor's share of overall fluctuations is high, then it is further evidence that nonlinearities play an important role in the credit cycle. The last column shows this variance decomposition for the second-order factor. Its median estimate is around 10%, indicating a secondary, although still quantitatively relevant, role. Similar to the results before, its credible set is wide.

H.4 Difference Between Positive and Negative Shocks

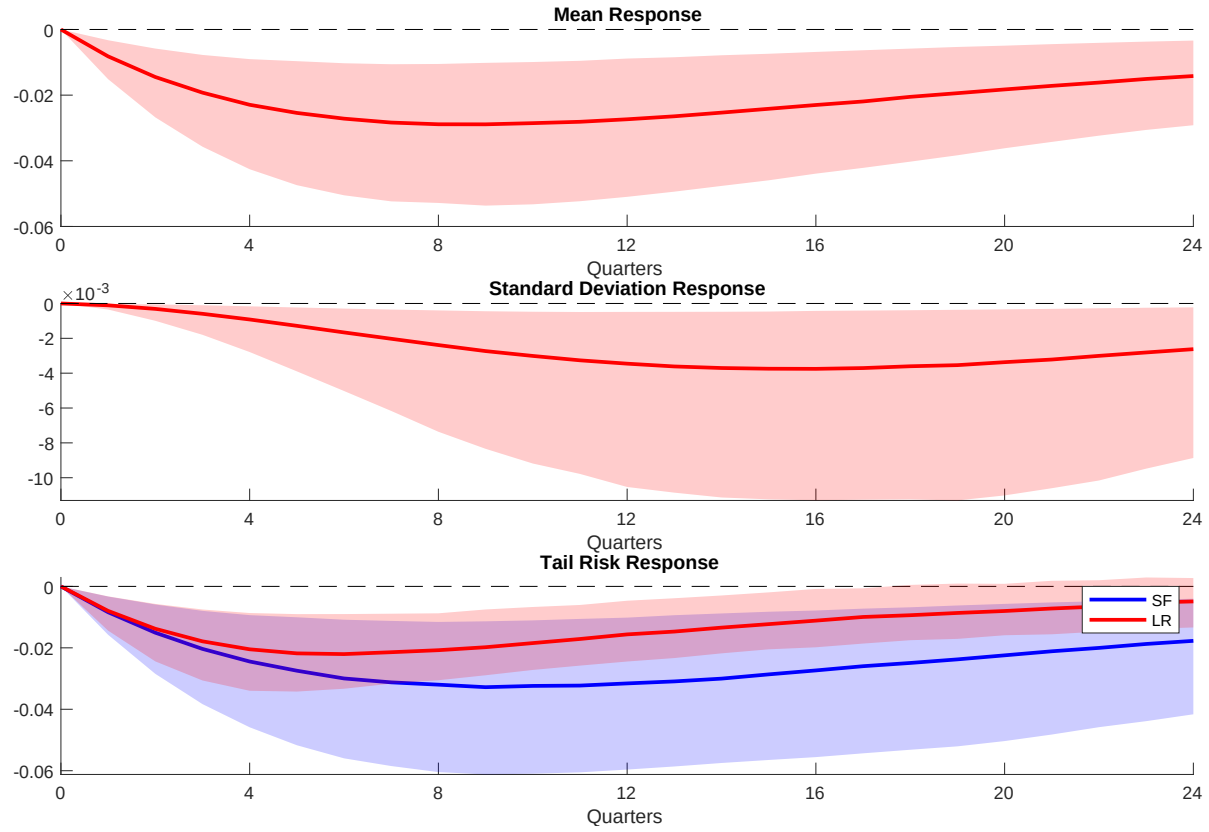


Figure 20: Credit Boom State (mid-2000s): Draw-by-draw differences between positive and negative shocks on the mean, standard deviation, and tail risk responses. Shaded areas denote 68% credible sets. SF is shortfall and LR is longrise.

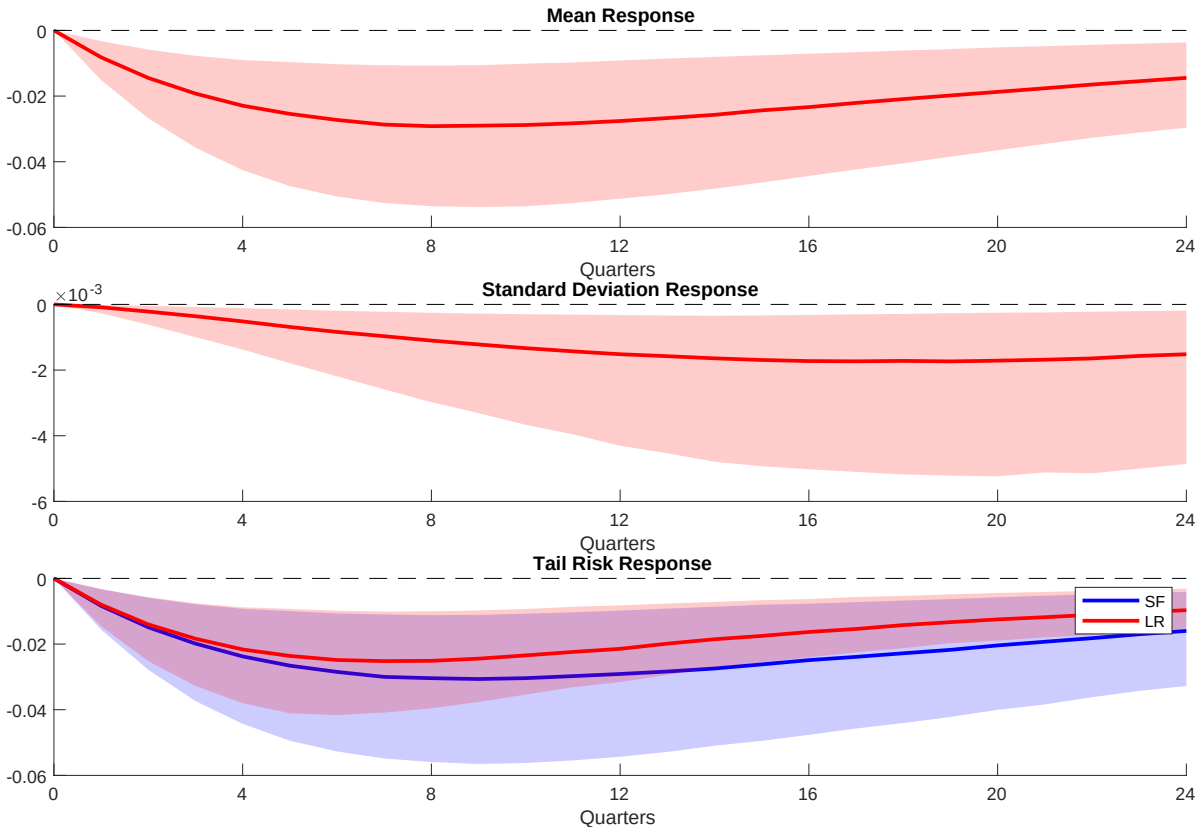


Figure 21: Credit Crunch State (2010): Draw-by-draw differences between positive and negative shocks on the mean, standard deviation, and tail risk responses. Shaded areas denote 68% credible sets. SF is shortfall and LR is longrise.

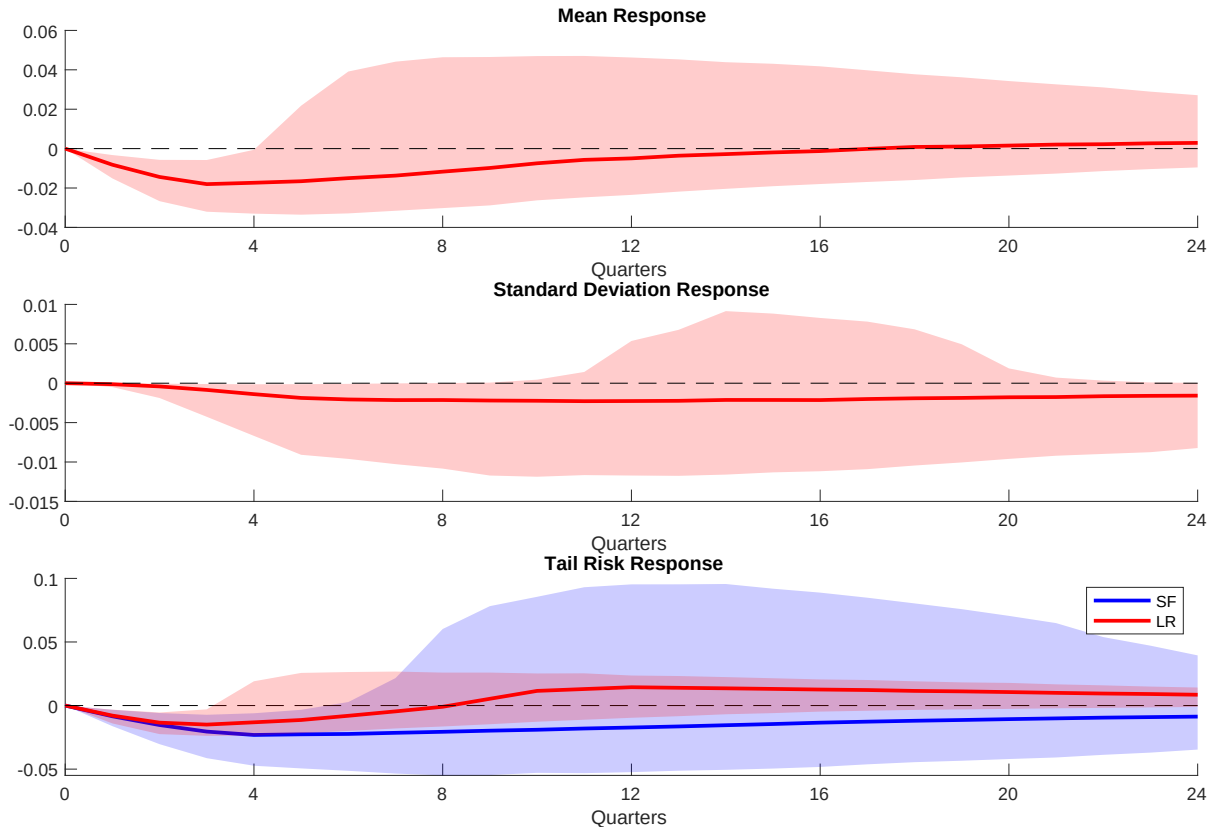


Figure 22: Credit Mix State (before the early 1990s recession): Draw-by-draw differences between positive and negative shocks on the mean, standard deviation, and tail risk responses. Shaded areas denote 68% credible sets. SF is shortfall and LR is longrise.

H.5 Parameter Estimates

Table 4: Parameter Estimates

	Prior	NLDF
h_x	$N(0.5, 1)$	0.922
		(0.901 , 0.938)
h_{xx}	$N(0, 5)$	-0.130
		(-0.223 , -0.053)
σ^2	$IW(v = 4, \eta = 1)$	0.062
		(0.050 , 0.077)
G_1	$N(0, 5)$	1.000
		(1.000 , 1.000)
G_2	$N(0, 5)$	1.318
		(1.162 , 1.470)
G_3	$N(0, 5)$	0.983
		(0.881 , 1.100)
G_4	$N(0, 5)$	-0.220
		(-0.327 , -0.116)
η_1^2	$IW\ (v = 4, \eta = 1)$	0.626
		(0.551 , 0.698)
η_2^2	$IW\ (v = 4, \eta = 1)$	0.282
		(0.234 , 0.347)
η_3^2	$IW\ (v = 4, \eta = 1)$	0.592
		(0.537 , 0.651)
η_4^2	$IW\ (v = 4, \eta = 1)$	0.968
		(0.888 , 1.05)

Median values of the posterior are reported. 16% and 84% are shown in brackets. NLDF is nonlinear dynamic factor model.

H.6 Comparison with Markov-Switching Dynamic Factor Model

We estimate the following two-state Markov-Switching dynamic factor model (MSDF):

$$\begin{aligned} y_t &= Gf_t + \eta\epsilon_t \\ f_t &= c(s_t) + h_x(s_t)f_{t-1} + \sigma(s_t)\nu_t \\ s_t|s_{t-1} &\sim P = \begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix} \end{aligned} \tag{38}$$

where ϵ_t is a 4×1 vector of iid $N(0, I_4)$ innovations, ν_t are iid $N(0,1)$ innovations, and s_t is the Markov state which switches according to the transition matrix P. We allow for $c(s_t)$, $h_x(s_t)$, and $\sigma(s_t)$ to switch. Additionally, we restrict $c(s_t)$ such that its unconditional mean is 0.

We use Bayesian methods to estimate the model. To compute the likelihood, we use the filter provided by Kim (1994). Table 5 shows the prior distributions for the parameters and estimates at the posterior mode.

The first state characterizes normal times with dynamics similar to the linear component of the NLDF model. The autoregressive component is high while the variance of the innovations is low, which generates a smoothly-varying and persistent latent factor. The second state is a crisis state that captures periods of volatile and fast-declining credit growth such as during the GFC. In this state, the intercept term $c(2)$ turns highly negative while the variance of the innovations $\sigma^2(2)$ increases by nearly tenfold compared to the normal state.

Table 5: Parameter Estimates: Markov-Switching Dynamic Factor Model

	Prior	Normal State	Crisis State
c	$N(0,2)$	0.015	-0.466
h_x	$N(0.5,1)$	0.910	0.080
σ^2	$IG\left(\alpha=2,\beta=0.5\right)$	0.044	0.399
G_1	$N(0,5)$	1.000	-
G_2	$N(0,5)$	1.653	-
G_3	$N(0,5)$	1.002	-
G_4	$N(0,5)$	-0.123	-
η_1^2	$IG\left(\alpha=2,\beta=0.5\right)$	0.746	-
η_2^2	$IG\left(\alpha=2,\beta=0.5\right)$	0.124	-
η_3^2	$IG\left(\alpha=2,\beta=0.5\right)$	0.665	-
η_4^2	$IG\left(\alpha=2,\beta=0.5\right)$	0.974	-
p_{11}	$B(0.75,0.5)$	0.990	-
p_{22}	$B(0.75,0.5)$	0.686	-

Posterior mode values are reported. Parameters that do not switch are denoted by a "-" in the Crisis State column. $IG(\alpha,\beta)$ corresponds to an inverse Gamma distribution with parameters α and β. $Be(\mu,std)$ corresponds to a Beta distribution with mean μ and standard deviation std. Finally, $N(\mu,std)$ is a Normal distribution with mean μ and standard deviation std.

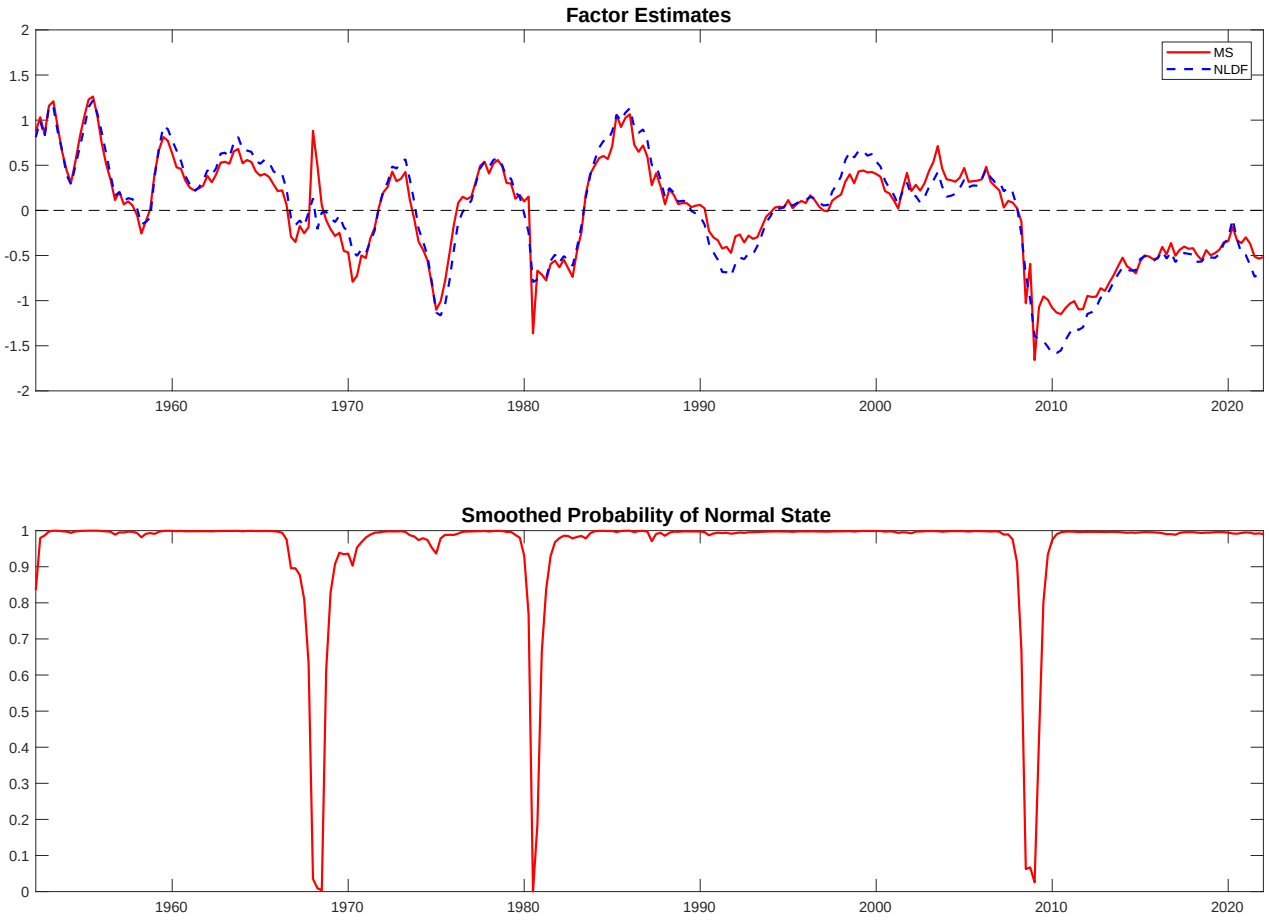
Figure 23 shows the smoothed estimates of the latent factor in the top panel and the smoothed probability of the normal state in the bottom panel. Comparing the red and dashed blue lines, we see that the factor estimates produced by the MSDF model are broadly similar to those from the NLDF model. There are differences, however, during times when the MSDF model estimates transitions into the crisis state. In those times, the factor estimates swing more wildly in the MSDF model and often show particularly steep declines within a quarter. Moreover, the NLDF model estimates a more sluggish recovery in the factor after the GFC compared to the MSDF model, which is consistent with the more persistent effects of shocks in the NLDF model during the credit crunch.

Figures 24 and 25 reproduce Figures 2 and 3, respectively, with the responses from the MSDF model overlaid. Recall that these figures show the effects of shocks to the latent factor. We use simulation methods to compute the impulse response functions from the MSDF model, so we take into account the possibility of future state transitions. Our definition of the IRFs is the same as what is presented in Section 3 of the main draft.[12]

There are three important points to note in Figure 24. First, state switches mainly change the

[12]We use $100,000$ simulations to compute the IRFs.

Figure 23: Estimated Credit Cycle and the Probability of the Normal State



NOTE: Smoothed factor estimates produced by the Markov-Switching model in red and the nonlinear dynamic factor model in dashed blue. The top panel shows the factor estimates and the bottom panel shows the smoothed probability of the normal state.

magnitudes of the impact effects of shocks, as can be seen by comparing the dashed green lines across the first and second columns. Upon moving from a normal state to a credit crunch, the effects of both positive and negative shocks approximately triple. Most of these changes are felt in the first few quarters, after which the propagation of the shocks behave quite similarly. Importantly, the MSDF model does not capture the hump-shaped response to shocks that the NLDF model does during the credit crunch. Second, as can be seen in the bottom panel of Figure 23, the MSDF model usually only estimates a high probability of being in the normal state or the crisis state. Since Markov-Switching is the only nonlinearity in the system, this means that the IRFs of shocks to the

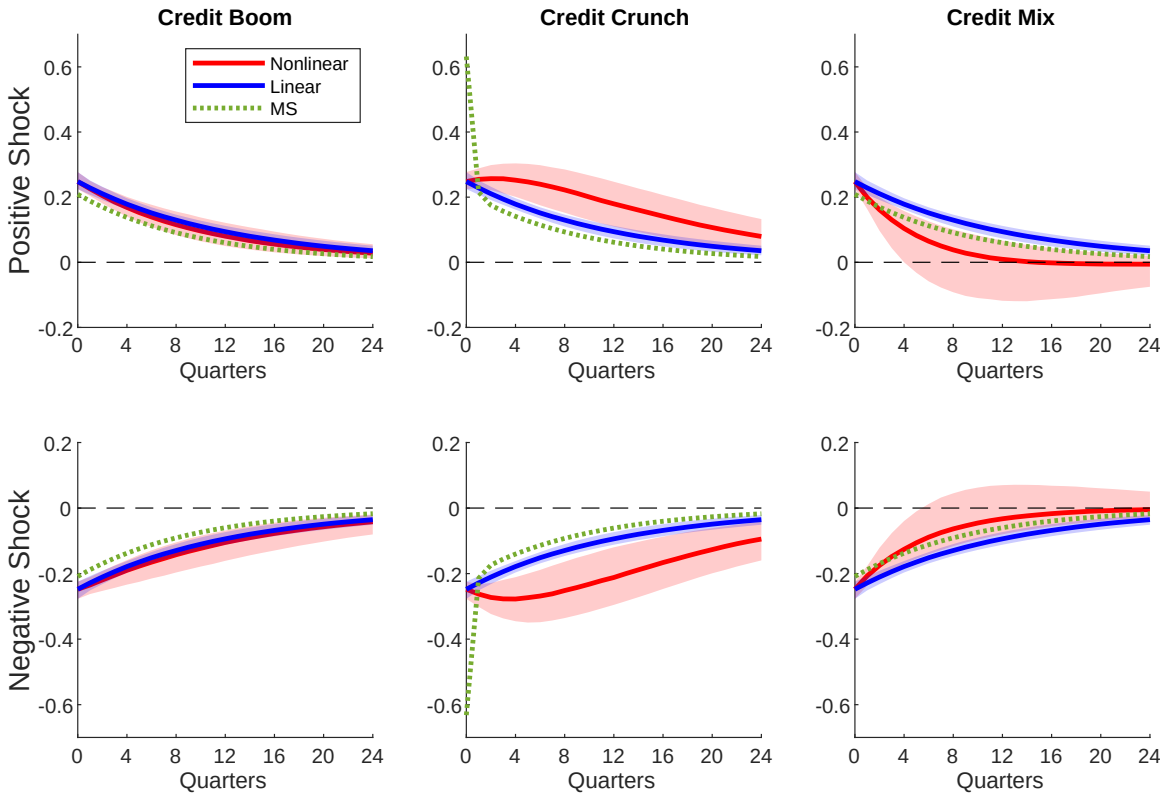
factor either look like those from the first column or the second. For instance, in the credit mix time period shown in the third column, the MSDF model produces IRFs that are almost identical to those from the credit boom. This is because in both time periods, the model estimates a probability near 1 of being in the normal state. Finally, comparing the responses from the top row to those from the bottom, we see that the MSDF model produces symmetric responses from positive versus negative shocks.

Figure 25 shows the responses of higher-order moments to shocks in the MSDF model during the GFC credit crunch. The main takeaway from this figure is that in the MSDF model there is no tight correlation between mean and volatility following shocks that generates a more volatile shortfall than longrise. Unlike in the NLDF model, there is no channel through which mean and volatility are connected following shocks to the factor in the MSDF model. Instead, any movements in the standard deviations of the distributions result from an interaction between future state switches and the levels of the factor. During the credit crunch, the standard deviations of the resulting conditional distributions increase following both positive and negative shocks, meaning that the longrise moves more than the shortfall after positive shocks and the shortfall moves more than the longrise after negative shocks.

In the MSDF model, the asymmetric tail risk dynamics primarily come from state switches. A transition to the crisis state simultaneously lowers the mean and increases volatility, thereby inducing a negative correlation between the conditional mean and volatility of the innovations and greatly shifting the lower tail of the distribution.

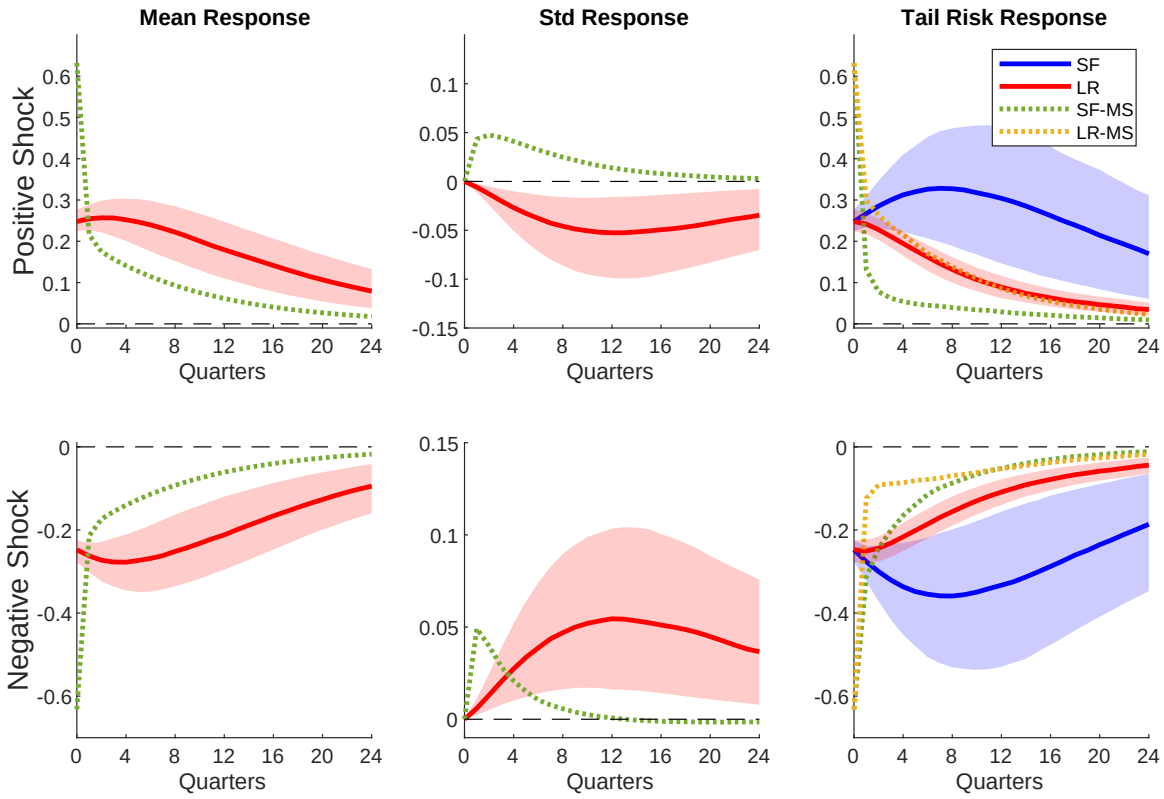
Compared to the NLDF model, the MSDF model does not capture the smoothly-varying state dependence of the factor dynamics nor the hump-shaped responses of the factor to innovations during credit crunches. While both models can produce the mean and volatility comovement that generates a volatile lower tail of the conditional distribution, they do so in different ways. The factor innovations do so in the NLDF model while the state switch does so in the MSDF model. We view the two models as complementary approaches to modeling nonlinearities.

Figure 24: State-Dependent Impulse Response Functions in Three Periods



NOTE: The red lines denote the responses of the overall factor following a positive shock (top row) and negative shock (bottom row), while the blue lines denote the responses of the linear component of the model. The first column conditions on a credit boom period in the mid-2000s, the second column conditions on a credit bust period in 2010, and the third column conditions on a mixed case before the early 1990s recession. The shaded areas denote 68% credible sets. The dashed green lines denote the responses of the Markov-Switching model. For the credit crunch time period, we condition on 2008:Q4, which is the quarter in the GFC with the highest probability of being in the crisis state. This does not coincide with the date used for the NLDF model because by then, the Markov-Switching model has already switched back to the normal state.

Figure 25: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk During the Credit Crunch

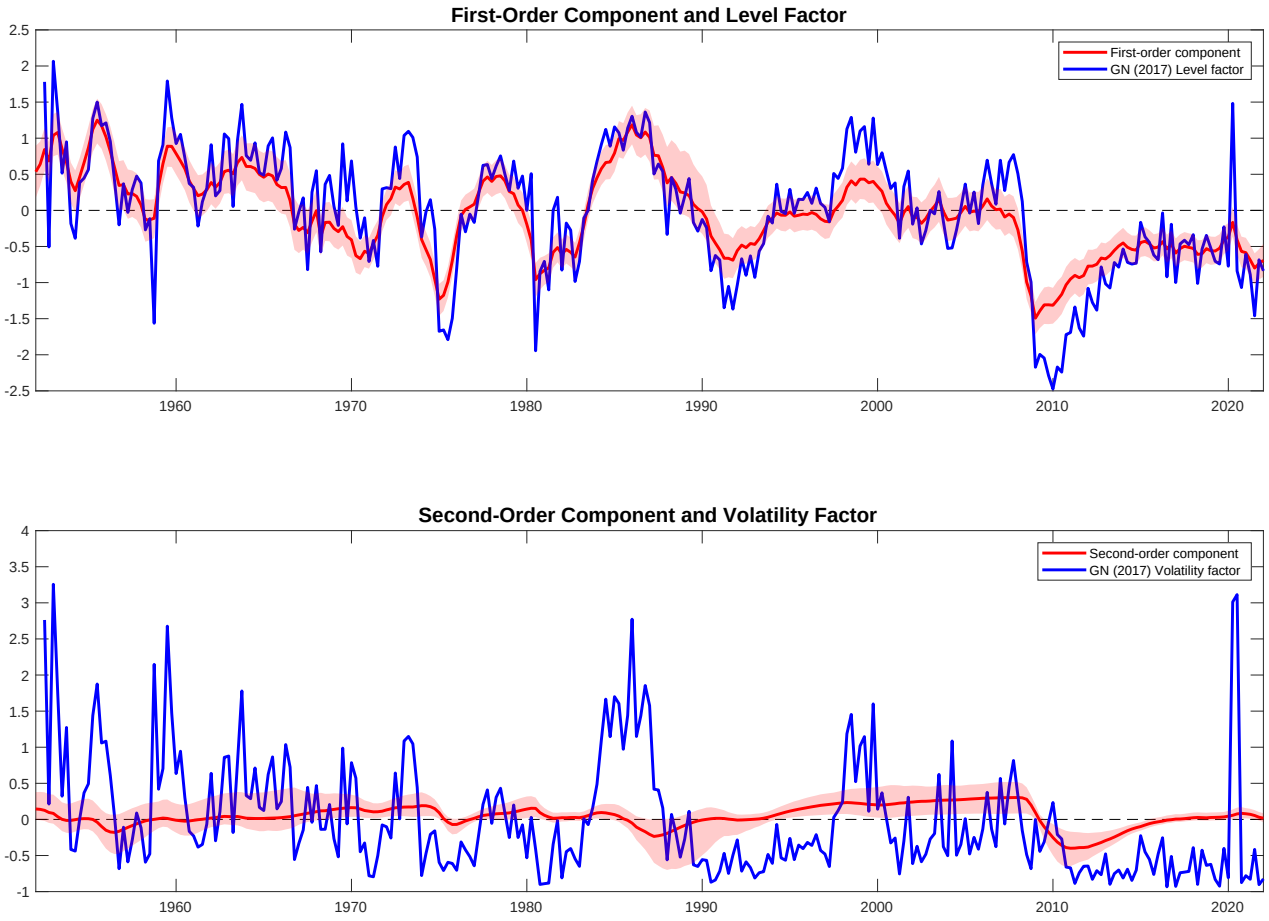


NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the demeaned overall factor produced by the nonlinear dynamic factor model during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. In the third column, the blue lines denote the shortfall (SF) response, while the red lines denote the longrise (LR) response. The shaded areas are 68% credible sets. The responses from the Markov-Switching model are shown in green in the first two columns. In the third column, the shortfall response is in green while the longrise response is in yellow. For the credit crunch time period, we condition on 2008:Q4, which is the quarter in the GFC with the highest probability of being in the crisis state. This does not coincide with the date used for the NLDF model because by then, the Markov-Switching model has already switched back to the normal state.

H.7 Comparison to Gorodnichenko and Ng (2017)

Gorodnichenko and Ng (2017) estimate two factors using FRED-MD data, which consists of 134 macroeconomic series. The first factor, level, is extracted using the original (normalized) series, while the second factor, volatility, uses the squared (normalized) series. We employ their methodology to extract factors from our credit growth data. Figure 26 compares the first- and second-order components of our one factor nonlinear model with those of the Gorodnichenko-Ng approach. The upper panel shows that our first-order component correlates strongly with the level factor. In contrast, our second-order component differs significantly from the volatility factor. This is visually clear during the late 1980s and the years following the 2008 financial crisis.

Figure 26: Comparing NLDF and Gorodnichenko and Ng (2017) Estimates



NOTE: Smoothed first- and second-order component estimates (red lines) produced by the nonlinear dynamic factor model compared to the linear and volatility factors produced using the methodology of Gorodnichenko and Ng (2017) (blue lines). The top panel shows the first-order component and level factor and the bottom panel shows the second-order component and volatility factor. The shaded areas denote 68% credible sets.

To compare goodness of fit, we examine the variances of the residuals from separate regressions of the four credit growth variables on extracted factors using Gorodnichenko and Ng (2017) and one and two factor versions of our NLDF model. Our regressions are of the form in Equation 39

$$y_t = c + \beta x_t + \epsilon_t, \ \epsilon_t \sim N(0, \sigma^2). \tag{39}$$

The x_t variables change depending on the regressions specification, with

$$x_t = \begin{cases} \text{level, volatility} & \text{if} \quad \text{Gorodnichenko and Ng (2017)} \\ f_t & \text{if} \quad \text{NLDF one factor} \\ f_{1,t}, f_{2,t} & \text{if} \quad \text{NLDF two factors.} \end{cases}$$

In the NLDF one factor case, f_t is the nonlinear factor from our one factor model while in the NLDF two factor case, $f_{1,t}$ is the first nonlinear factor and $f_{2,t}$ is the second from the two factor version of our model discussed in Section H.10.

The results are shown in Table 6. Compared to the one factor version of the model, the Gorodnichenko and Ng (2017) methodology produces a better fit for all variables except for household credit growth. This is not surprising since Gorodnichenko and Ng (2017) extracts two factors compared to one factor in our baseline case. When we move to the NLDF with two factors, our model fits the data better for nonfinancial business and household credit growth while doing worse for financial and government credit growth. Overall, we conclude that our NLDF model is complementary to Gorodnichenko and Ng (2017)'s framework.

Table 6: Model Fit Comparison: Variance of Regression Residuals

Data	GN	NLDF	NLDF Two Factors
Nonfinancial business	0.35	0.55	0.15
Household	0.50	0.27	0.10
Financial	1.36	2.63	2.71
Government	1.73	2.46	2.25

The second column of the table shows the variances of the residuals from the regressions using the Gorodnichenko and Ng (2017) factors. The third column uses our one factor NLDF model. The fourth column uses our two factor NLDF model.

H.8 Persistent Measurement Error

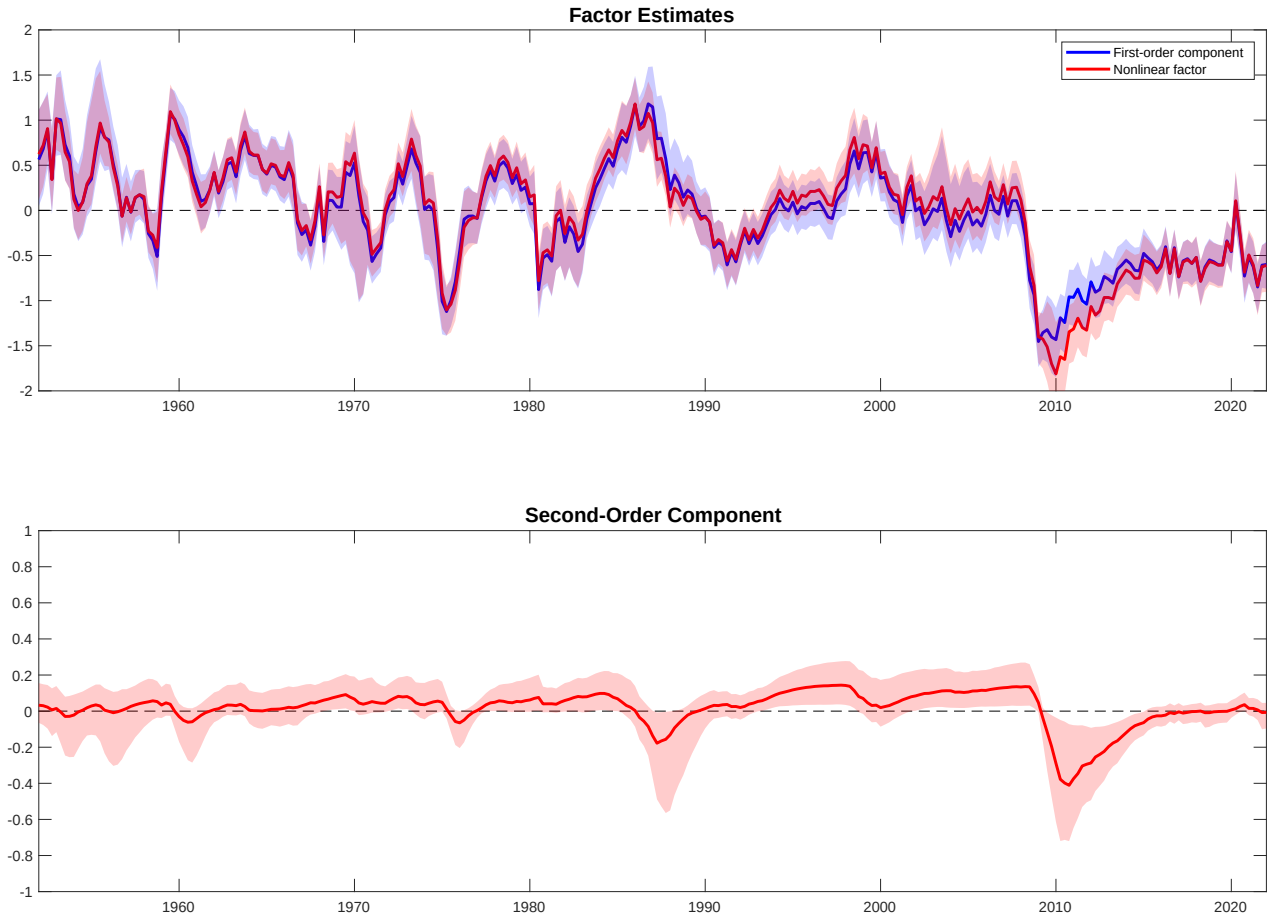
We examine the robustness of our results to persistent measurement error. To this end, we extend our baseline NLDF model to allow for the measurement error of each observable variable i to have persistence ρ_i. The extended model is shown in Equation 40.

$$\begin{aligned} y_t &= Gf_t + \epsilon_t \\ f_t &= c + f_t^f + f_t^s \\ f_t^f &= h_x f_{t-1}^f + \sigma\nu_t,\ \nu_t \sim N(0,1) \\ f_t^s &= h_x f_{t-1}^s + \frac{1}{2}h_{xx}\left(f_{t-1}^f\right)^2 \\ \epsilon_{i,t} &= \rho_i \epsilon_{i,t-1} + \eta_i \xi_{i,t},\ \xi_{i,t} \sim N(0,1),\ \forall i = 1, ..., N \end{aligned} \tag{40}$$

We continue to model a single factor and set $G_1 = 1$ for identification. The estimation specifications for this model are the same as for the benchmark estimation.

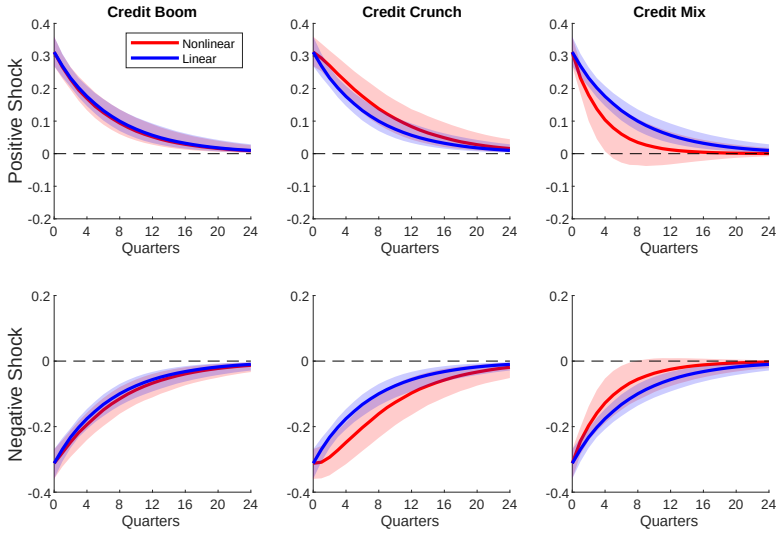
The inclusion of persistent measurement errors does not change the results. Figure 27 shows the smoothed factors, which look similar to those of our main results. Figure 28 illustrates the state-dependent nature of the IRFs and the response of the nonlinear factor is still close to that of the linear component during the credit boom, amplified during the credit crunch, and dampened during the mixed case period. Finally, Figure 29 shows the higher-order moment responses during the credit crunch. A positive shock increases the mean while decreasing the volatility of the conditional distribution, which raises the shortfall more than the longrise. A negative shock decreases the mean and increases the volatility, leading to a larger decline in the shortfall compared the longrise. These results are again comparable to those of the benchmark model.

Figure 27: Estimated Credit Cycle and the Contribution of the Second-Order Factor



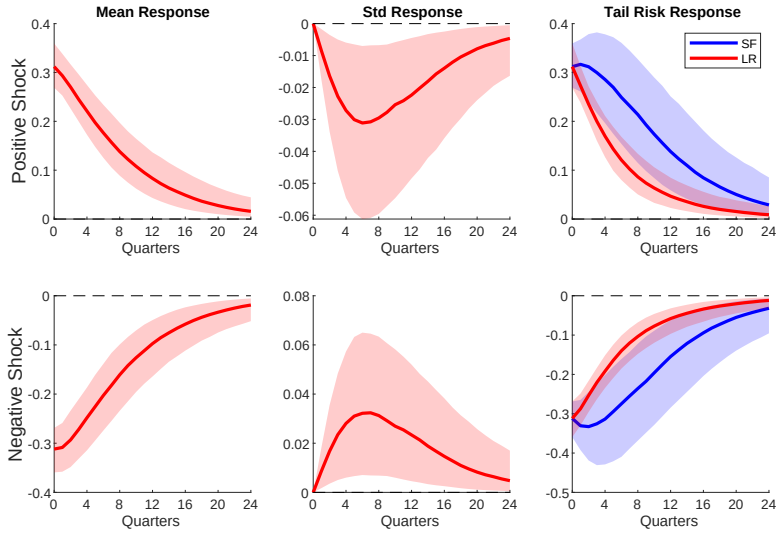
NOTE: Smoothed factor estimates produced by the nonlinear dynamic factor model. The top panel shows the demeaned factor estimates, with the red line being the estimate of the nonlinear factor and the blue line the estimate of the first-order component. The shaded areas denote 68% credible sets. The bottom panel shows the demeaned second-order component with 68% credible sets.

Figure 28: State-Dependent Impulse Response Functions in Three Periods



NOTE: The red lines denote the responses of the overall factor following a positive shock (top row) and negative shock (bottom row), while the blue lines denote the responses of the linear component of the model. The first column conditions on a credit boom period in the mid-2000s, the second column conditions on a credit bust period in 2010, and the third column conditions on a mixed case before the early 1990s recession. The shaded areas denote 68% credible sets.

Figure 29: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk During the Credit Crunch



NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the demeaned overall factor produced by the nonlinear dynamic factor model during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. In the third column, the blue lines denote the shortfall (SF) response, while the red lines denote the longrise (LR) response. The shaded areas are 68% credible sets.

Table 7 shows the parameter estimates from the model. The persistence of the measurement error for nonfinancial business credit and government credit growth are positive and are the only variables that have 68% credible sets that do not contain 0. Although the posterior median of the persistence for household credit growth is greater than 0.6, its credible set is wide and contains 0, indicating a high degree of estimation uncertainty.

Table 7: Parameter Estimates: Persistent Measurement Error

	Prior	NLDF
h_x	$N(0.5, 1)$	0.867
		(0.823 , 0.908)
h_{xx}	$N(0, 5)$	-0.120
		(-0.241 , -0.030)
σ^2	$IW(v = 4, \eta = 1)$	0.098
		(0.073 , 0.13)
G_1	$N(0, 5)$	1.000
		(1.000 , 1.000)
G_2	$N(0, 5)$	1.014
		(0.807 , 1.376)
G_3	$N(0, 5)$	1.085
		(0.883 , 1.281)
G_4	$N(0, 5)$	-0.132
		(-0.307 , 0.065)
ρ_1	$N(0, 1)$	0.630
		(0.559 , 0.704)
ρ_2	$N(0, 1)$	0.607
		(-0.253 , 0.674)
ρ_3	$N(0, 1)$	-0.082
		(-0.201 , 0.214)
ρ_4	$N(0, 1)$	0.464
		(0.405 , 0.519)
η_1^2	$IW\ (v = 4, \eta = 1)$	0.382
		(0.341 , 0.426)
η_2^2	$IW\ (v = 4, \eta = 1)$	0.311
		(0.135 , 0.356)
η_3^2	$IW\ (v = 4, \eta = 1)$	0.458
		(0.384 , 0.639)
η_4^2	$IW\ (v = 4, \eta = 1)$	0.764
		(0.703 , 0.831)

Median values of the posterior are reported. 16% and 84% are shown in brackets.

H.9 Separate Weights for Second-Order Component

We consider an extension of our baseline NLDF model where the first- and second-order components have different weights in the measurement equation. The measurement equation is modified as follows:

$$y_t = G_1 f_t^f + G_2(c + f_t^s) + \eta\epsilon_t,\ \epsilon_t \sim N(0, I).$$

We introduce a new factor loading G_2 that determines how the observable loads onto the second-order component. As discussed in Section A.1, the separate loading for the second-order component introduces additional identification issues. To identify the model, we set $G_{1,1} = 1$ and $G_{1,2} = 1$ for nonfinancial business credit growth. We impose that the persistence of the first- and second-order components are the same. The estimation specifications for this model are the same as for the benchmark estimation.

Table 8: Parameter Estimates: Different Weights

	Prior	NLDF
h_x	$N(0.5, 1)$	0.867
		(0.845 , 0.886)
h_{xx}	$N(0, 5)$	-2.132
		(-2.644 , -1.725)
σ^2	$IW(v = 4, \eta = 1)$	0.062
		(0.052 , 0.074)
$G_{1,1}$	$N(0, 5)$	1.000
		(1.000 , 1.000)
$G_{2,1}$	$N(0, 5)$	-0.603
		(-0.749 , -0.468)
$G_{3,1}$	$N(0, 5)$	-0.037
		(-0.176 , 0.093)
$G_{4,1}$	$N(0, 5)$	0.021
		(-0.122 , 0.166)
$G_{1,2}$	$N(0, 5)$	1.000
		(1.000 , 1.000)
$G_{2,2}$	$N(0, 5)$	0.633
		(0.573 , 0.697)
$G_{3,2}$	$N(0, 5)$	0.666
		(0.603 , 0.729)
$G_{4,2}$	$N(0, 5)$	-0.261
		(-0.334 , -0.185)
η_1^2	$IW\ (v = 4, \eta = 1)$	0.233
		(0.208 , 0.263)
η_2^2	$IW\ (v = 4, \eta = 1)$	0.410
		(0.373 , 0.452)
η_3^2	$IW\ (v = 4, \eta = 1)$	0.593
		(0.543 , 0.650)
η_4^2	$IW\ (v = 4, \eta = 1)$	0.927
		(0.856 , 1.01)

Median values of the posterior are reported. 16% and 84% are shown in brackets. NLDF is nonlinear dynamic factor model.

H.10 Two Factors

We consider a two-factor version of the NLDF model. For parsimony, we restrict each factor to have independent dynamics. This means that the autoregressive h_x matrix and the square root of the variance covariance matrix of the shocks σ are diagonal. Also, the h_{xx} matrix is such that the second-order components of the factors only load onto their own respective squared first-order components. This assumption allows both factors to have nonlinear dynamics while restricting the possible interactions in the second-order components for parsimony. Equation 41 shows the model that we estimate.

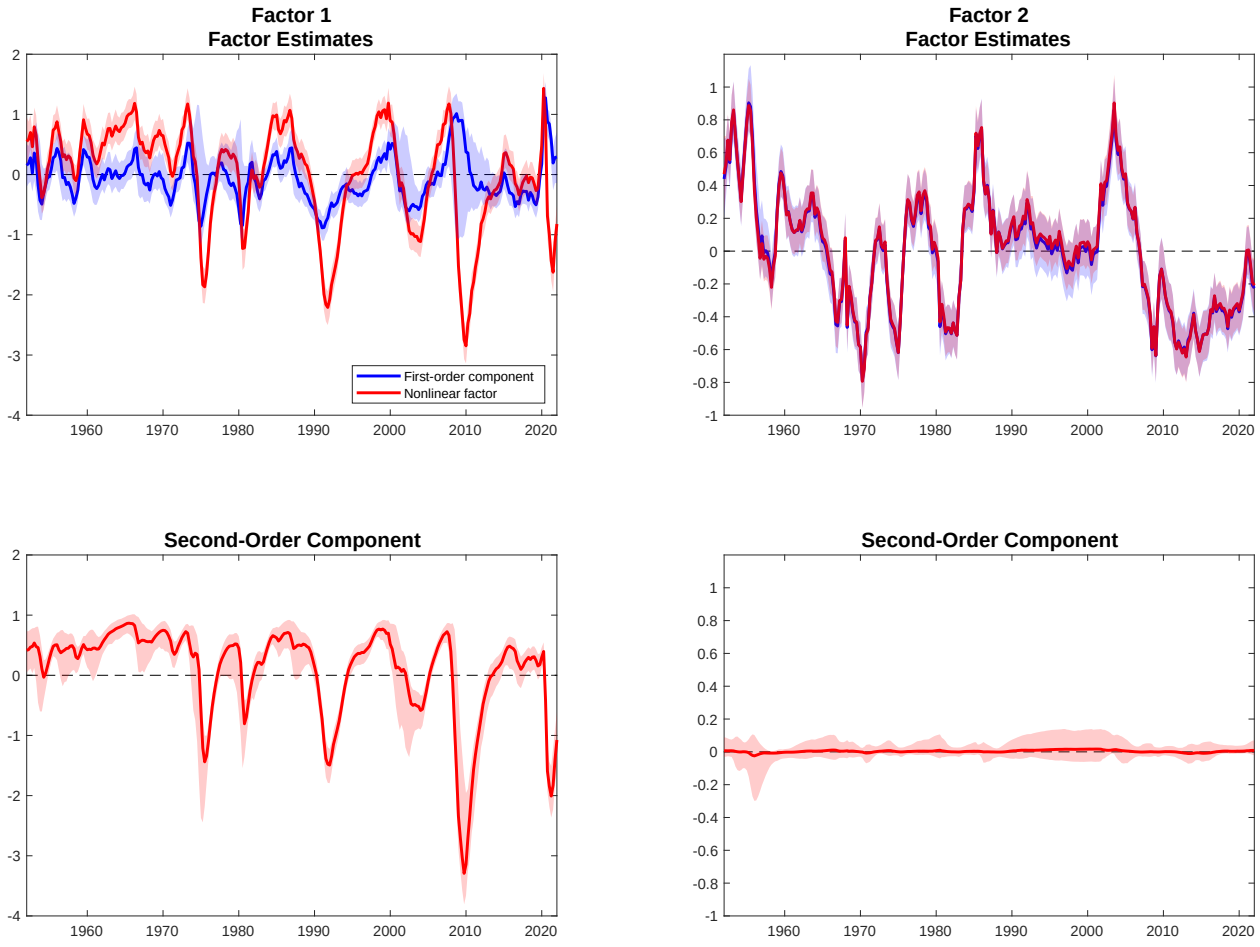
$$
\begin{aligned}
y_t &= G \begin{pmatrix} f_{1,t} \\ f_{2,t} \end{pmatrix} + \begin{pmatrix} \eta_1 & 0 \\ 0 & \eta_2 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, I_2 \right) \\
\begin{pmatrix} f^f_{1,t} \\ f^f_{2,t} \end{pmatrix} &= \begin{pmatrix} h_{x,1} & 0 \\ 0 & h_{x,2} \end{pmatrix} \begin{pmatrix} f^f_{1,t-1} \\ f^f_{2,t-1} \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, I_2 \right) \\
\begin{pmatrix} f^s_{1,t} \\ f^s_{2,t} \end{pmatrix} &= \begin{pmatrix} h_{x,1} & 0 \\ 0 & h_{x,2} \end{pmatrix} \begin{pmatrix} f^s_{1,t-1} \\ f^s_{2,t-1} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} h_{xx,1} & 0 & 0 \\ 0 & 0 & h_{xx,2} \end{pmatrix} \begin{pmatrix} \left(f^f_{1,t-1}\right)^2 \\ f^f_{1,t-1} f^f_{2,t-1} \\ \left(f^f_{2,t-1}\right)^2 \end{pmatrix}
\end{aligned}
\tag{41}
$$

We address the identification issues discussed in Apepndix Section A.1 by imposing a 2×2 lower triangular matrix on the factor loadings for nonfinancial business credit and government credit growth, with nonfinancial business credit growth having a unit loading on the first factor and 0 loading on the second factor and government credit growth having a unit loading on the second factor. The estimation specifications for this model are the same as for the benchmark estimation.

Figure 30 shows the smoothed factors in the first row and their second-order components in the second row. The first factor has important nonlinear dynamics, as evidenced by the large fluctuations of the second-order component. In contrast, the second factor is nearly linear with a negligible contribution from the second-order component. Table 9 shows the parameter estimates from the

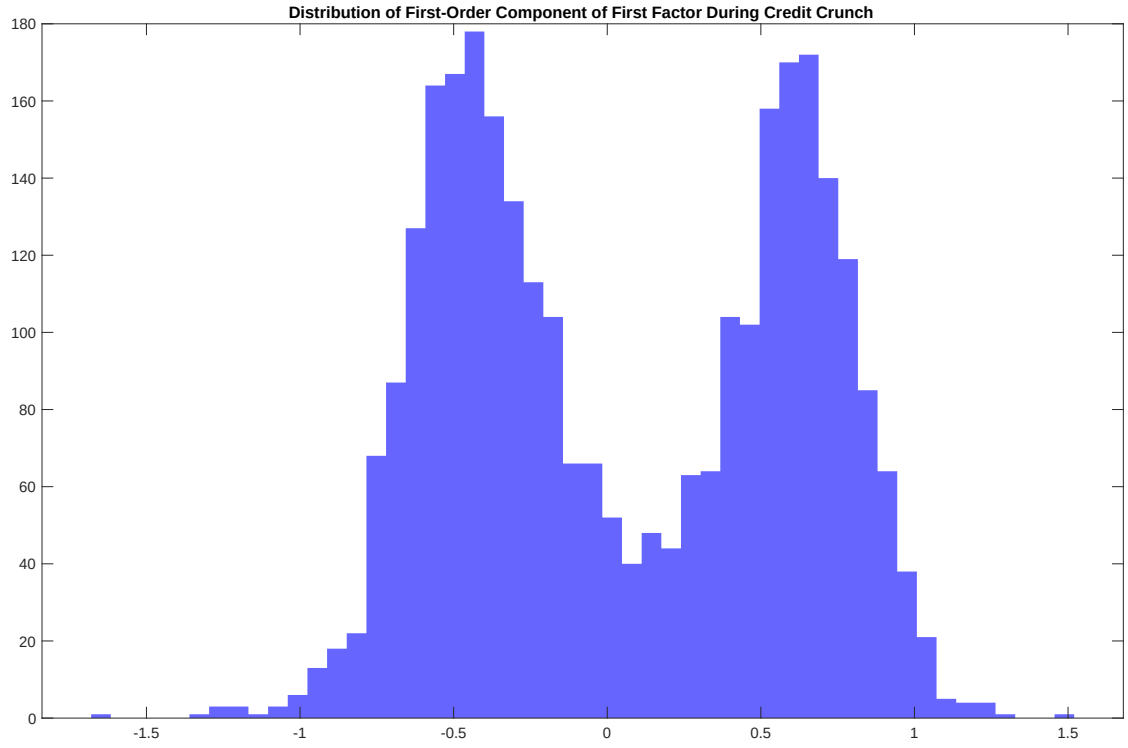
model. The first factor plays a similar role as the nonlinear factor from the baseline model, loading positively on all of the credit growth variables except government credit growth, which has a negative loading. The second factor does not load on nonfinancial business credit growth by construction and loads positively on the other credit growth series. It plays a secondary role overall for credit growth fluctuations.

Figure 30: Estimated Factors and the Contributions of the Second-Order Factor



NOTE: Smoothed factor estimates produced by the nonlinear dynamic factor model. The top row shows the demeaned factor estimates, with the red line being the estimate of the nonlinear factor and the blue line the estimate of the first-order component. The shaded areas denote 68% credible sets. The bottom row shows the demeaned second-order components with 68% credible sets.

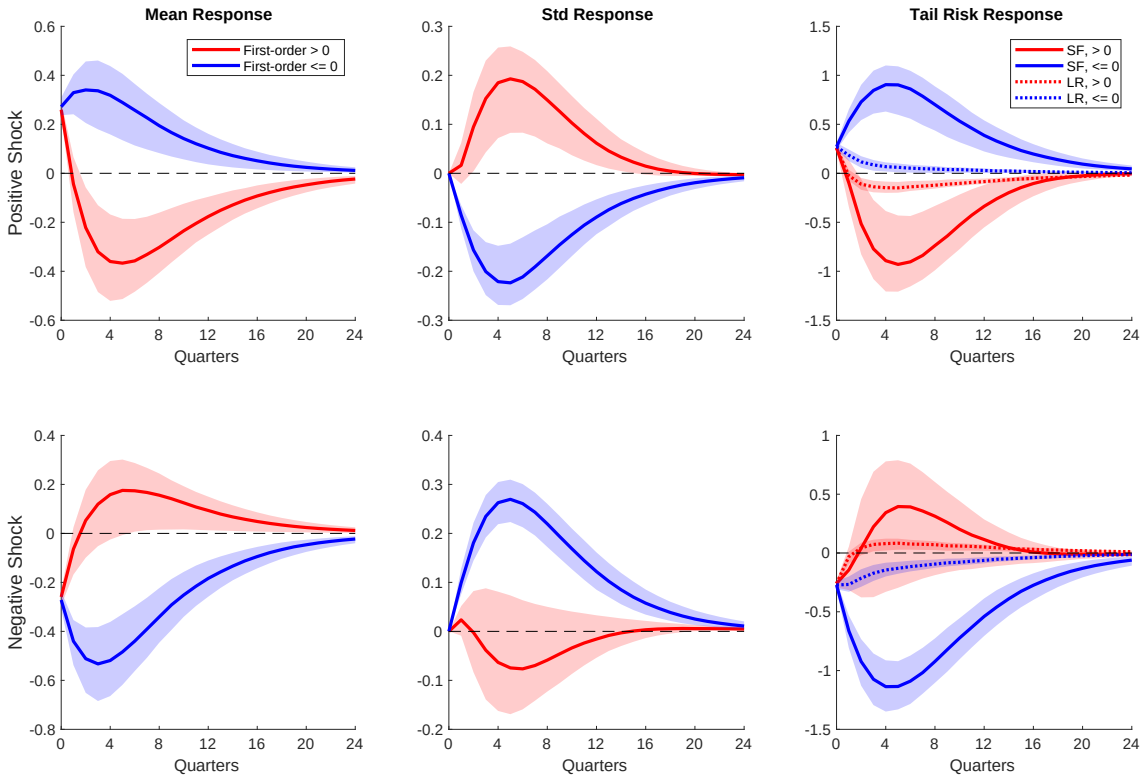
Figure 31: Bimodality of First-Order Component of First Factor



NOTE: This figure shows the histogram of posterior draws from the first-order component of the first factor during the credit crunch (2010:Q4).

The fluctuations of the first factor are dominated by its second-order component, whereas the primary purpose of the first-order component is to drive the second-order component. This dynamic leads to situations, such as around the credit crunch in 2010, where the first-order component is imprecisely estimated around 0. As the blue credible sets in the first panel of Figure 30 show, the smoothed distribution of the first-order component widens during this time as positive and negative values become similarly likely. The posterior distribution of the smoothed first-order component shows a clear bimodality, as seen in Figure 31. This happens because depending on whether the first-order component is positive or negative, different configurations of shocks lead to similar factor dynamics moving forward that fit the data equally well. This similarity means that a positive versus negative first-order component becomes more difficult to distinguish from each other.

Figure 32: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk of the First Factor During the Credit Crunch



NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the demeaned first factor produced by the nonlinear dynamic factor model during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. The red lines denote instances in which the linear component of the first factor is positive, while the blue lines denote instances in which the linear component is negative. In the third column, the solid lines denote the shortfall (SF) response, while the dashed lines denote the longrise (LR) response. The shaded areas are 68% credible sets.

Figure 32 further illustrates this fact by plotting the mean, standard deviation, and tail risk responses of the first factor during the credit crunch. The top row shows the effects of a positive shock while the bottom row shows the effects of a negative shock. We separate the responses into the two regimes that are shown in Figure 31: the blue lines denote the responses conditional on the first-order component being below 0 while the red lines denote the responses conditional on the first-order component being above 0. This conditioning has important implications for the responses to shocks. The shaded areas are 68% credible bands conditional on being in each regime.

When the first-order component is positive, a positive shock generates an increase in the factor initially, but it quickly reverses and becomes persistently negative as the second-order component

dominates (red lines in the top row). The standard deviation increases and the shortfall and longrise both decrease, with the shortfall decreasing by more than the longrise. When the first-order component is negative, a positive shock raises the factor and decreases its standard deviation (blue lines in the top row). The combination of these two effects increases the shortfall by more than the longrise. The bottom row shows the effects of a negative shock. When the first-order component is negative (blue lines in the bottom row), the overall factor immediately decreases and stays depressed, while the standard deviation increases. The shortfall declines by more than the longrise. When the first-order component is positive, the effects of the shock are milder (red lines in the bottom row).

Table 9: Parameter Estimates: Two Factors

	Prior	NLDF
$h_{x,1}$	$N(0.5,1)$	0.837
		(0.816 , 0.858)
$h_{xx,1}$	$N(0,5)$	-1.656
		(-2.225 , -1.162)
σ_1^2	$IW(v=4,\eta=1)$	0.070
		(0.055 , 0.090)
$h_{x,2}$	$N(0.5,1)$	0.888
		(0.851 , 0.917)
$h_{xx,2}$	$N(0,5)$	-0.029
		(-0.169 , 0.108)
σ_2^2	$IW(v=4,\eta=1)$	0.036
		(0.029 , 0.044)
$G_{1,1}$	$N(0,5)$	1.000
		(1.000 , 1.000)
$G_{2,1}$	$N(0,5)$	0.442
		(0.338 , 0.550)
$G_{3,1}$	$N(0,5)$	0.598
		(0.526 , 0.674)
$G_{4,1}$	$N(0,5)$	-0.340
		(-0.429 , -0.244)
$G_{1,2}$	$N(0,5)$	0.000
		(0.000 , 0.000)
$G_{2,2}$	$N(0,5)$	2.029
		(1.819 , 2.267)
$G_{3,2}$	$N(0,5)$	0.798
		(0.643 , 0.978)
$G_{4,2}$	$N(0,5)$	1.000
		(1.000 , 1.000)
η_1^2	$IW\ (v=4,\eta=1)$	0.234
		(0.206 , 0.268)
η_2^2	$IW\ (v=4,\eta=1)$	0.167
		(0.143 , 0.193)
η_3^2	$IW\ (v=4,\eta=1)$	0.591
		(0.540 , 0.645)
η_4^2	$IW\ (v=4,\eta=1)$	0.944
		(0.863 , 1.03)

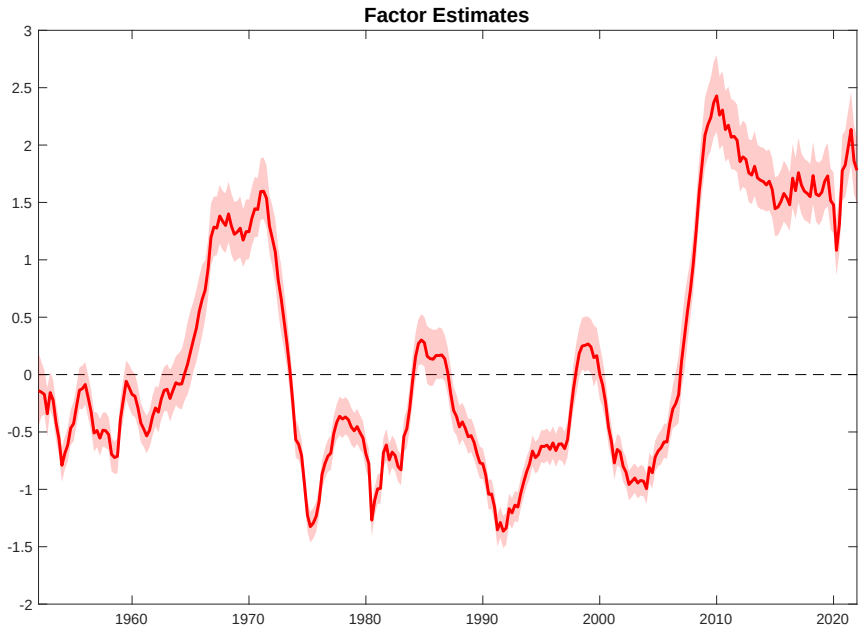
Median values of the posterior are reported. 16% and 84% are shown in brackets. NLDF is nonlinear dynamic factor model.

H.11 Quadratic Measurement Equation

We estimate a one factor version of the model with quadratic measurement equation presented in Appendix Section G. We take $300,000$ draws from the posterior distribution with $120,000$ burn-in. We save every 60th draw to form our posterior distributions of interest.

Figure 33 shows the estimated linear factor driving the system. The factor is highly persistent with extended peaks in the late 1960s into the early 1970s and during and after the GFC. The time around the GFC is especially interesting. In the data, the three private credit growth series collapsed, yet the model estimates a large increase in the factor. This is counterintuitive if we only focus on the effects of the level of the factor, because the private credit growth series are estimated to load positively on it (Table 10). However, when the factor becomes especially large in magnitude, the effects of the squared factor dominate those from the level of the factor. The private credit growth series load negatively on the squared factor and this channel leads to declines in those series from increases in the factor during the GFC. Therefore, the factor parses the rise and subsequent fall in private credit growth as a consequence of an ever-increasing factor. When the factor initially increases, the effects from the rise in the level dominate, and private credit growth rises. As the factor continues increasing, the squared factor dominates and credit growth declines.

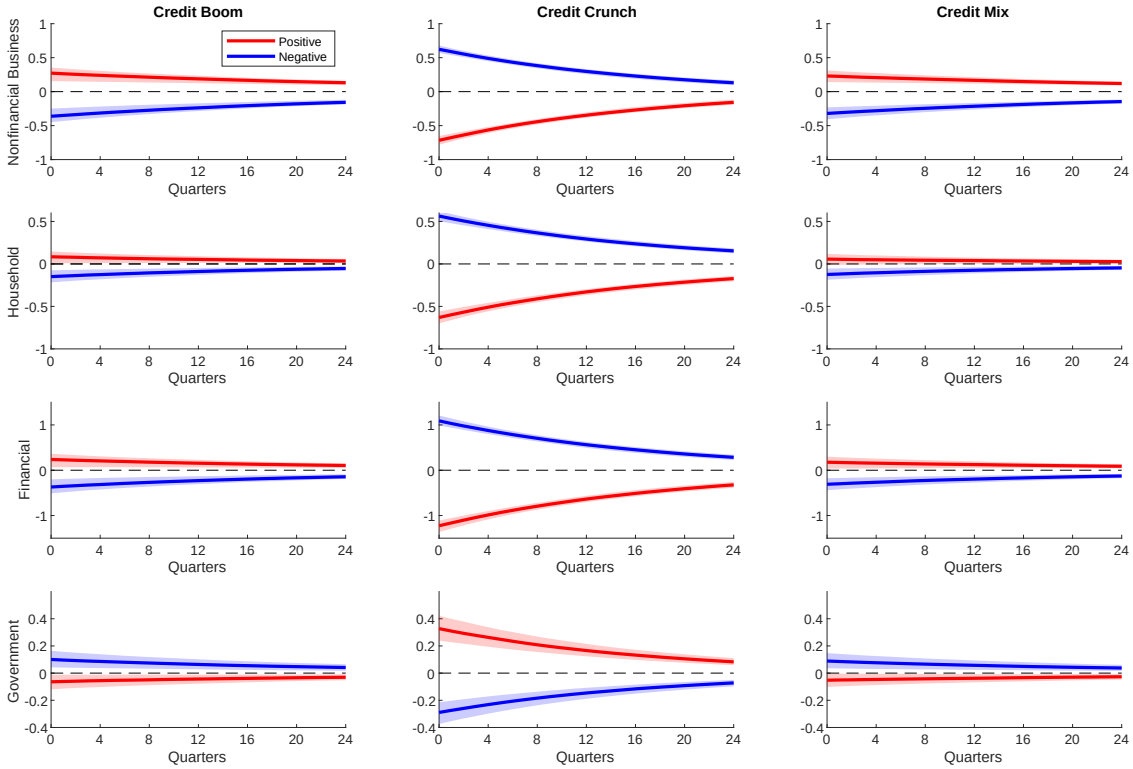
Figure 33: Estimated Factor



NOTE: Smoothed factor estimates produced by the dynamic factor model with quadratic measurement equation. The shaded areas denote 68% credible sets.

Figure 34 provides further support for this result by presenting the responses of the credit growth series to positive (red) and negative (blue) shocks to the factor conditioned on the same three time periods as in the main paper. During the credit boom shown in the first column, increases in the factor raise nonfinancial business, household, and financial credit growth while lowering government credit growth, as would be expected if the level of the factor dominates. As we move to the credit crunch in the second column, the direction of the responses flip. A positive shock to the factor now decreases the private credit growth series and increases government credit growth. This flip is consistent with the quadratic component of the model dominating. The magnitude of the responses, whether to a positive or negative shock, also greatly increase, consistent with the more volatile movements in credit growth during the crunch. Finally, the credit mix time shown in the third column shows similar responses as the credit boom, suggesting that this model does not distinguish across these two periods.

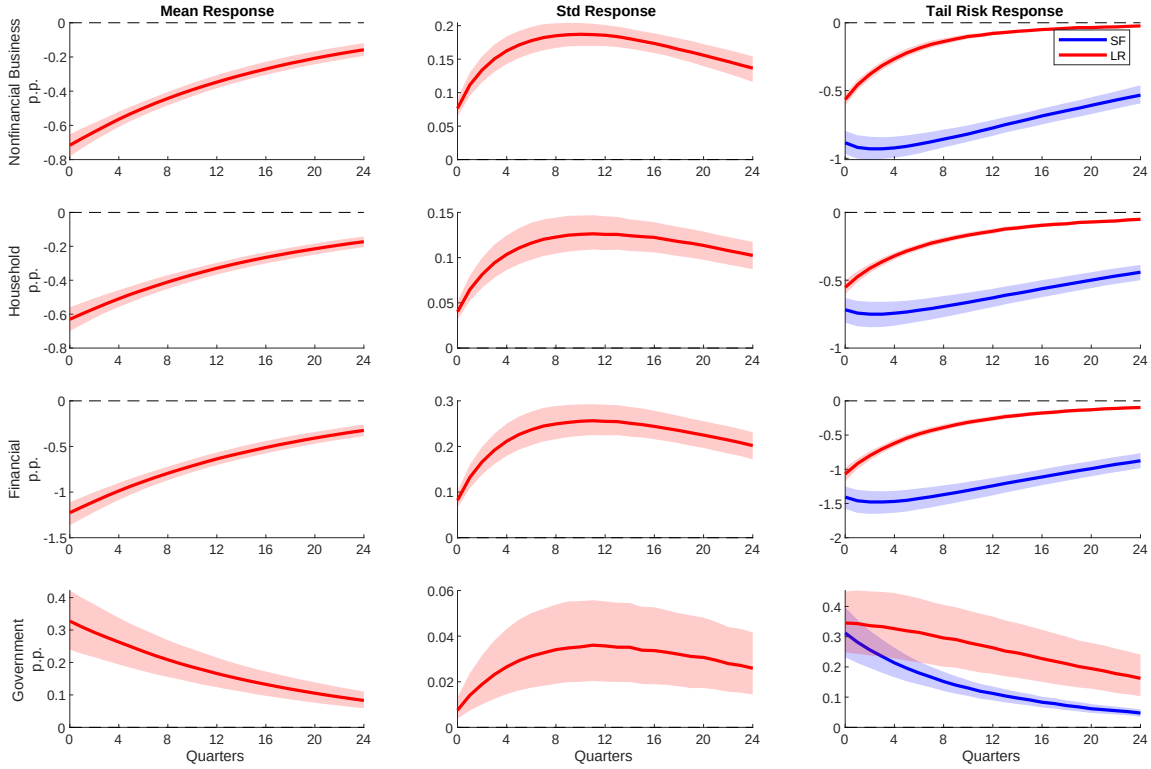
Figure 34: State-Dependent Impulse Response Functions in Three Periods



NOTE: The red lines denote the responses of the credit growth variables following a positive shock and the blue lines denote the responses following negative shock. The first column conditions on a credit boom period in the mid-2000s, the second column conditions on a credit bust period in 2010, and the third column conditions on a mixed case before the early 1990s recession. The first row shows the responses of nonfinancial business credit, the second row household credit, the third row financial credit, and the fourth row government credit. The shaded areas denote 68% credible sets.

Figure 35 shows the mean, standard deviation, and tail risk responses across horizons for the credit growth series during the credit crunch. A shock that lowers the means of the private credit growth series increases the standard deviations. The combined effects of these responses leads to a larger decline in the shortfall compared to the longrise of the distributions. The effects are long-lasting and hump-shaped for the standard deviation responses because of the high persistence of the shock. For government credit growth, the shock increases the mean and standard deviation together, which raises the longrise by more than the shortfall.

Figure 35: Impulse Response Functions of the Mean, Standard Deviation, and Tail Risk During the Credit Crunch



NOTE: Impulse response functions of the mean, standard deviation, and 5% shortfall and longrise of the credit growth variables during the credit crunch period in 2010. The responses to a positive shock are shown in the top row, and the responses to a negative shock are shown in the bottom row. In the third column, the blue lines denote the shortfall (SF) response, while the red lines denote the longrise (LR) response. The shaded areas are 68% credible sets.

Table 10: Parameter Estimates: Quadratic

	Prior	Quadratic
h_x	$N(0.5, 1)$	0.975
		(0.971 , 0.979)
σ^2	$IW(v = 4, \eta = 1)$	0.048
		(0.039 , 0.058)
$G_{1,1}$	$N(0, 5)$	1.000
		(1.000 , 1.000)
$G_{2,1}$	$N(0, 5)$	0.173
		(0.081 , 0.262)
$G_{3,1}$	$N(0, 5)$	0.308
		(0.223 , 0.389)
$G_{4,1}$	$N(0, 5)$	-0.138
		(-0.228 , -0.046)
$G_{1,2}$	$N(0, 5)$	-1.002
		(-1.247 , -0.809)
$G_{2,2}$	$N(0, 5)$	-0.621
		(-0.789 , -0.485)
$G_{3,2}$	$N(0, 5)$	-0.655
		(-0.828 , -0.517)
$G_{4,2}$	$N(0, 5)$	0.245
		(0.166 , 0.344)
η_1^2	$IW\ (v = 4, \eta = 1)$	0.235
		(0.206 , 0.270)
η_2^2	$IW\ (v = 4, \eta = 1)$	0.524
		(0.476 , 0.575)
η_3^2	$IW\ (v = 4, \eta = 1)$	0.568
		(0.518 , 0.617)
η_4^2	$IW\ (v = 4, \eta = 1)$	0.931
		(0.855 , 1.010)

Median values of the posterior are reported. 16% and 84% are shown in brackets. Quadratic is the model with a quadratic measurement equation.

I Interest Rate Factor

I.1 Benchmark Estimation

In this section, we illustrate that our methodology can also accommodate nonlinear measurement equations. We suggest a parsimonious model of a forward interest rate factor that imposes ELB restrictions on the forward rates in the measurement equation. Additionally, we allow for our second-order factor dynamics. As we only specify one factor, this exercise is meant to be illustrative of possible extensions of our modeling framework.

Empirical Set-up. We estimate our model on the first differences of one-month-forward rate data at 3-month, 6-month, 1-year, 2-year, 3-year, 5-year, and 10-year maturities from February 1990 to September 2019. We follow the approach of Wu and Xia (2016) in constructing the one-month forward rates for seven maturities based on the nominal yield curve data from Gurkaynak et al. (2007). We use the code provided by Wu and Xia (2016). The end-of-month monthly data spans the period from January 1990 to September 2019, and the maturities used are the same as in the original paper: 3 and 6 months and 1, 2, 5, 7, and 10 years.

Our choice to model the forward rates in first differences instead of levels follows the recommendation of Onatski and Wang (2021). The ELB is set at 0.3% on the series. We consider the following model of differences in forward rates:

$$\Delta forward_t^h = m_h + \begin{cases} G_h(c + f_t^f + f_t^s) + \eta^h \epsilon_t^h & \text{if } \widehat{S}_t^h >= 0.3 \\ -m_h + \eta^h \epsilon_t^h & \text{otherwise} \end{cases} \tag{42}$$

where $c = -\frac{1}{2}\frac{h_{xx}\sigma^2}{(1-h_x)(1-h_x^2)}$, $\widehat{S}_t^h = \sum_{\tau=2}^{t}\left(m_h + G_h\left(c + f_\tau^f + f_\tau^s\right)\right) + forward_1^h$, $\Delta forward_t^h = forward_t^h - forward_{t-1}^h$, and index h stands for the maturity. We model the latent factor according to our second-order dynamics described in Equation 3 and normalize the factor loading G_1 on the three-month rate to be 1. We also allow for maturity-specific constants m_h that capture any differences in the average forward rate changes across maturities. Our model is an econometric model of forward curve dynamics, so we do not impose arbitrage-free restrictions.

The model features two key nonlinearities. First, there is an ELB constraint in the measurement equation, which removes any influence of the factor if it predicts a rate lower than 0.3% in levels. Instead, the measurement equation is set to $\Delta forward_t^h = \eta^h \epsilon_t^h$. At the ELB, the observed change in the forward rate should be 0 p.p., as the level is stuck at 0.3%. The measurement error picks up a residual difference. This constraint leads to a regime-switching behavior and plays a similar role to the ELB constraint in other yield curve models that model rates in levels. Second, the latent factor dynamics are allowed to be nonlinear.

We estimate the nonlinear model using Metropolis Hastings combined with the bootstrap particle filter. We use $500,000$ particles in the particle filter. We take $80,000$ draws from the posterior distribution after burn-in. We construct the posterior distributions for our results by taking every 40th parameter draw. Further estimation details can be found in Appendix Section E.

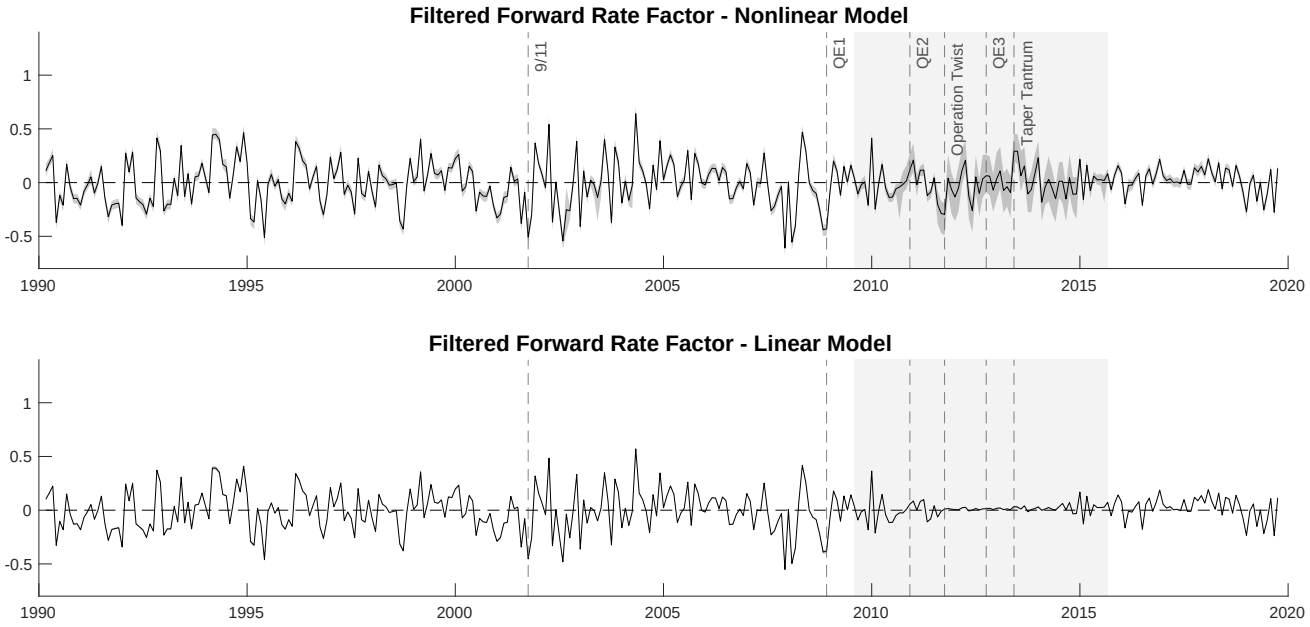
As a comparison, we also estimate a linear version of the model in which we do not model the ELB in the measurement equation and impose linear factor dynamics. Insofar as the nonlinearities we consider are empirically important, the nonlinear model should fit the data better than the linear model does. If the h_{xx} term in the nonlinear model is estimated to be insignificant, then we interpret that as evidence against state-dependent dynamics at the ELB.

Historical Estimates of the Forward Rate Factor. We first discuss the historical filtered estimates of the forward rate factor implied by our nonlinear model and compare it with a linear model that does not account for the ELB restriction and imposes only linear factor dynamics. The top panel in Figure 36 shows the filtered estimates of the factor produced by our nonlinear model. The factor captures the contours of historical forward curve movements. For example, we catch the rapid drop in short- and long-term rates in the early 1990s, which was then followed by a tightening cycle in monetary policy during the middle of the decade. From 1995 through 1996, the 10-year forward rate dropped more than 2 percentage points, which our factor captures. As we move on to the 2000s, our estimated factor reflects the easing cycle in the early 2000s followed by tightening beginning in 2004. The factor then rapidly drops entering into the GFC. In the middle of the GFC, short-term interest rates hit their ELB, which is highlighted by the gray rectangular area in the figure. Because our model can account for an ELB in the measurement equation, how-

ever, it still estimates variability in the factor, chiefly the continued decline in long-term rates. The lack of variation of shorter-term maturities leads to a widening of the uncertainty in the estimated factor as more forward rates hit their ELB. Finally, as we exit the ELB, the factor captures the rise in short-term forward rates.

It is worth emphasizing that our factor matches some important events in recent monetary policy (the dashed vertical lines in Figure 36). For example, there is a sharp decline in the factor in the aftermath of 9/11. A similar sudden drop is observed around the announcement of the Fed's first round of quantitative easing (QE1). In contrast, the Taper Tantrum of 2013 coincides with a rise in the factor. Interestingly, our estimated factor shows that QE3 did not result in a change of the stance of monetary policy.

Figure 36: Filtered Estimates of the Forward Rate Factor from the Nonlinear and Linear Models



NOTE: Filtered estimates of the forward rate factor estimated by the nonlinear model (top panel) and linear model (bottom panel) and 80% credible sets (shaded area). The rectangular shaded areas denote the periods in which the 3-month forward rate is at the effective lower bound. QE1, QE2, and QE3 are the first, second, and third rounds of quantitative easing by the Fed.

The bottom panel shows the corresponding filtered estimates from a linear model. Outside of the shaded ELB period, the two models estimate similar factors, foreshadowing the limited role played by the second-order factor dynamics. During the ELB period, however, the factor estimates

diverge. The linear model is constrained by the absence of variation in short-term forward rates, as they are stuck at 0.3%, while long-term rates continue to vary. Indeed, these fluctuations in the longer maturity rates inform the dynamics of our nonlinear factor during the zero lower bound episodes. On balance, the linear model estimates little variation during the ELB period, thereby sacrificing fit to the long-term yields.

Importance of the Nonlinear Components. How important are the nonlinear additions to the model? From the filtered estimates, we see clear evidence that the ELB restriction tangibly changes the factor estimates. Moreover, a marginal likelihood comparison between the two models using the modified harmonic mean (Geweke, 1999) and a truncation parameter of 0.95 shows that the nonlinear model is heavily favored by the data at 687 versus 539 log points. These results are robust to other values of the truncation parameter such as 0.75 and 0.5. Taken together, they suggest that allowing for nonlinearities is important to understand forward curve dynamics.

The central question in our investigation is whether the ELB produced structural changes in forward curve dynamics and therefore changes in the behavior of longer-term forward rates unconstrained by the ELB. This question can be answered by examining whether the statistical gains from the nonlinear model are primarily due to the ELB constraint on the measurement equation, the second-order factor dynamics, or a combination of the two. Table 11 shows the 80% credible sets of parameter estimates. There, we can see that the credible sets for h_{xx} – the key parameter that governs the second-order factor – ranges from -0.183 to 0.638. These estimates contain 0 and suggest that higher-order factor dynamics play a limited role. We can also estimate a version of the model in which we maintain the ELB restriction in Equation 42, but we impose linear factor dynamics. This version of the model produces filtered factors similar to the fully nonlinear model and fits the data slightly better in a marginal likelihood sense (689 versus 687 log points).

Our empirical evidence then is in favor of the idea that the ELB mainly was a restriction on the behavior of short-term forward rates. There is little evidence of nonlinearities in the factor dynamics, at least using our model. Therefore, dynamics of the factor continued to propagate linearly as in unconstrained times.

I.2 Tailoring the Estimation to the Effective Lower Bound

We discuss here how we modify the estimation presented in Appendix Section E.2 to account for the nonlinear measurement equation in Equation 42, reproduced below for convenience.

$$\Delta forward_t^h = m_h + \begin{cases} G_h(c + f_t^f + f_t^s) + \eta^h \epsilon_t^h & \text{if } \widehat{S}_t^h >= 0.3 \\ -m_h + \eta^h \epsilon_t^h & \text{otherwise} \end{cases} \tag{43}$$

where $c = -\frac{1}{2}\frac{h_{xx}\sigma^2}{(1-h_x)(1-h_x^2)}$, $\widehat{S}_t^h = \sum_{\tau=2}^{t}\left(m_h + G_h\left(c + f_\tau^f + f_\tau^s\right)\right) + forward_1^h$, $\Delta forward_t^h = forward_t^h - forward_{t-1}^h$, and index h stands for the maturity. We model the latent factor according to our second-order dynamics:

$$\begin{aligned} f_t^f &= h_x f_{t-1}^f + \sigma\nu_t \\ f_t^s &= h_x f_{t-1}^s + \frac{1}{2}h_{xx}\left(f_{t-1}^f\right)^2. \end{aligned} \tag{44}$$

Relative to the benchmark NLDF with a linear measurement equation, there are three main differences. First, we have forward-rate-specific constants m_h that capture the long-run mean of each series. This is a straightforward addition to the Metropolis Hastings algorithm, and we add a block to the estimation procedure. Second, we have to keep track of $\widehat{S}_t$ in the particle filter, which is the sum of the entire path of the particle. To account for this sum, we add an additional component to the particle called

$$\widehat{S}_t^{(j)} = \left(c + f_t^{f,(j)} + f_t^{s,(j)}\right) + \widehat{S}_{t-1}^{(j)}.$$

The conversion from $\widehat{S}_t^{(j)}$ to $\widehat{S}_t^{h,(j)}$ for each maturity h is straightforward from their respective formulas.

Third, the measurement equation that we use to evaluate the weight of the particle in the particle filter changes depending on whether $\widehat{S}_t^{h,(j)}$ is greater than or less than 0.3. This feature affects the prediction step in our algorithm.

I.3 Estimation of the Linear Model

We also estimate a linear version of the model that removes the nonlinearity in the measurement equation and only allows for first-order factor dynamics. We use the same Metropolis Hastings scheme as laid out in Appendix Section E.2, with two differences. First, we have the extra parameters m_h that we estimate as an additional block. Second, we use the Kalman Filter instead of the particle filter to estimate the model, as it is now a linear Gaussian state-space model.

I.4 Parameter Estimates

Table 11: Parameter Estimates

	Prior	NLDF and ELB	Linear factor and ELB	Linear
h_x	$N(0,1)$	0.144	0.170	0.190
		(0.098 , 0.190)	(0.108 , 0.231)	(0.121 , 0.265)
h_{xx}	$N(0,5)$	0.265	0.000	0.000
		(-0.183 , 0.638)	(0.000 , 0.000)	(0.000 , 0.000)
σ^2	$IW(v=4,\eta=1)$	0.053	0.055	0.030
		(0.046 , 0.061)	(0.047 , 0.063)	(0.026 , 0.034)
G_1	$N(0,5)$	1.000	1.000	1.000
		(1.000 , 1.000)	(1.000 , 1.000)	(1.000 , 1.000)
G_2	$N(0,5)$	1.142	1.114	1.302
		(1.069 , 1.224)	(1.057 , 1.193)	(1.237 , 1.385)
G_3	$N(0,5)$	1.356	1.347	1.654
		(1.294 , 1.425)	(1.288 , 1.421)	(1.577 , 1.741)
G_4	$N(0,5)$	1.397	1.379	1.724
		(1.332 , 1.473)	(1.308 , 1.459)	(1.627 , 1.836)
G_5	$N(0,5)$	0.932	0.921	1.113
		(0.867 , 1.044)	(0.851 , 1.013)	(1.004 , 1.224)
G_6	$N(0,5)$	0.733	0.737	0.803
		(0.665 , 0.817)	(0.670 , 0.801)	(0.690 , 0.920)
G_7	$N(0,5)$	0.538	0.575	0.558
		(0.452 , 0.616)	(0.505 , 0.643)	(0.0447 , 0.673)
η_1^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^1\right)\right)$	0.016	0.016	0.02
		(0.015 , 0.017)	(0.014 , 0.018)	(0.018 , 0.022)
η_2^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^2\right)\right)$	0.009	0.009	0.010
		(0.008 , 0.010)	(0.008 , 0.010)	(0.009 , 0.011)
η_3^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^3\right)\right)$	0.002	0.001	0.001
		(0.001 , 0.002)	(0.001 , 0.002)	(0.001 , 0.002)
η_4^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^4\right)\right)$	0.013	0.013	0.015
		(0.011 , 0.014)	(0.012 , 0.015)	(0.013 , 0.017)
η_5^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^5\right)\right)$	0.045	0.045	0.057
		(0.040 , 0.052)	(0.040 , 0.049)	(0.052 , 0.063)
η_6^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^6\right)\right)$	0.058	0.058	0.07
		(0.053 , 0.065)	(0.052 , 0.063)	(0.064 , 0.078)
η_7^2	$IW\left(v=4,\eta=\frac{1}{5}Var\left(\Delta forward^7\right)\right)$	0.064	0.063	0.071
		(0.058 , 0.070)	(0.057 , 0.070)	(0.065 , 0.079)

Median values of the posterior are reported. 10% and 90% are shown in brackets. ML is marginal likelihood computed with Geweke (1999) modified harmonic mean estimator with truncation parameter of 0.95. NLDF is nonlinear dynamic factor model. ELB is effective lower bound.

	Prior	NLDF and ELB	Linear factor and ELB	Linear
m_1	$N(-0.019, 0.004^2)$	-0.030	-0.031	-0.019
		(-0.033 , -0.028)	(-0.033 , -0.028)	(-0.023 , -0.014)
m_2	$N(-0.019, 0.004^2)$	-0.024	-0.023	-0.019
		(-0.026 , -0.021)	(-0.025 , -0.020)	(-0.023 , -0.015)
m_3	$N(-0.019, 0.004^2)$	-0.018	-0.018	-0.019
		(-0.020 , -0.015)	(-0.020 , -0.015)	(-0.023 , -0.015)
m_4	$N(-0.019, 0.004^2)$	-0.015	-0.016	-0.019
		(-0.019 , -0.013)	(-0.019 , -0.013)	(-0.024 , -0.015)
m_5	$N(-0.019, 0.004^2)$	-0.014	-0.014	-0.019
		(-0.016 , -0.010)	(-0.016 , -0.011)	(-0.024 , -0.014)
m_6	$N(-0.019, 0.004^2)$	-0.016	-0.016	-0.019
		(-0.019 , -0.012)	(-0.019 , -0.012)	(-0.024 , -0.014)
m_7	$N(-0.019, 0.004^2)$	-0.018	-0.019	-0.019
		(-0.021 , -0.015)	(-0.021 , -0.013)	(-0.024 , -0.014)
ML		687	689	539

Median values of the posterior are reported. 10% and 90% are shown in brackets. ML is marginal likelihood computed with Geweke (1999) modified harmonic mean estimator with truncation parameter of 0.95. NLDF is nonlinear dynamic factor model. ELB is effective lower bound.

I.5 Persistent Measurement Error

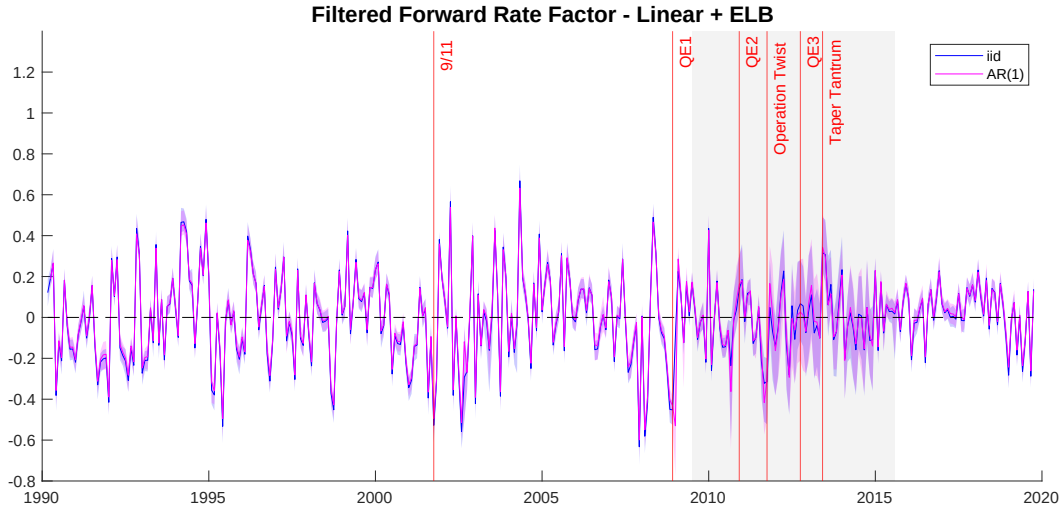
We examine the persistence of measurement errors in our model. To this end, we reformulate the measurement equation to have the form

$$\Delta forward_t^h = m_h + \begin{cases} G_h f_t + \eta^h \epsilon_t^h & \text{if } \widehat{S}_t^h >= 0.3 \\ -m_h + \eta^h \epsilon_t^h & \text{otherwise} \end{cases} \tag{45}$$

For each of the observables i, we assume that the autocorrelation in the residuals is AR(1): $\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \eta_{i,t}\varepsilon_{it}$, where $\varepsilon_{i,t} \sim iid\ N(0,1)$. We specify the factor as evolving linearly, as it is the preferred model in our main results. We then estimate this model using the same dataset and MCMC procedure as in the main paper. We use the same priors as in the benchmark model. We specify the priors on the measurement error persistence terms as $N(0,25)$.

Figure 37 compares factor estimates of the model with iid errors in blue with those from the model with AR(1) errors in magenta. The two estimates are nearly identical. This makes sense, as the observables we use are already in first differences and therefore not persistent.

Figure 37: Filtered Estimates of the Forward Rate Factor: Robustness to Persistent Measurement Error



NOTE: Filtered estimates of the forward rate factor estimated by Linear + ELB model with iid (blue) and AR(1) (magenta) measurement error along with 80% credible sets (shaded area). The gray shaded areas denote the periods in which the 3-month forward rate is at the effective lower bound. QE1, QE2, and QE3 are the first, second, and third rounds of quantitative easing by the Fed.

I.6 Quadratic Measurement Equation

Following Kim and Singleton (2012), we estimate a quadratic measurement equation model on the same dataset as in the benchmark results. This setup specifies the measurement equation as quadratic in the factors and while the factors propagate linearly. The model we estimate is:

$$\begin{aligned} \Delta forward_t^h &= m_h + G_{h,1} f_t + G_{h,2}(c + f_t^2) + \eta^h \epsilon_t^h \\ f_t &= h_x f_{t-1} + \sigma \nu_t \end{aligned} \tag{46}$$

where $c = -\frac{\sigma^2}{1-h_x^2}$, $\Delta forward_t^h = forward_t^h - forward_{t-1}^h$, and index h stands for the maturity. We use the same priors as in the benchmark model. We specify the priors on the quadratic factor loading coefficients as $N(0, 5)$.

Note that this model does not explicitly account for the ELB. Although there are papers that use a quadratic measurement equation to enforce the ELB, our specification does not naturally do so because we model the forward rates in first differences as opposed to levels. Nevertheless, it is still of interest to see how well a quadratic measurement equation can capture the ELB or other nonlinearities in forward rate behavior.

Table 12 shows the marginal likelihood from the model. The quadratic measurement equation model fits the data better than the linear model, but its fit is still dominated by the models that explicitly account for the ELB.

Table 12: Marginal Likelihood Comparison

Quadratic	544
NLDF and ELB	687
Linear factor and ELB	689
Linear	539

The marginal likelihood is computed with the Geweke (1999) modified harmonic mean estimator and truncation parameter of 0.95. Quadratic is the model with a quadratic measurement equation. NLDF is nonlinear dynamic factor model. ELB is effective lower bound.