State-Dependent Impulse Responses under Uncertainty

Seojin Jung*

October 7, 2025

Abstract

I study the finite-sample behavior of state-dependent impulse response estimators under uncertainty. I start from a design where the true data generating process and its transition rule are known, then add misspecifications step by step. The target impulse response is defined in a potential outcome framework, while estimation relies on prediction-based methods. Performance is evaluated by CRPS across single estimators and model-averaging schemes.

Simulations show that CRPS is mainly determined by the transition process and the shock size. As shocks grow, LP estimators tend to outperform, though truncated-lag VARs can remain competitive. Bayesian and plain or bias-corrected estimators often move in opposite directions. A simpler transition process can lower errors when the true rule is noisy, while imposing linearity on a nonlinear process can produce distortions.

In the empirical analysis with three inflation-related regimes, uncertainty shocks are state dependent and build more slowly in favorable states, whereas monetary shocks show little state dependence.

Keywords: state-dependence; impulse response; misspecification; potential outcomes; local projections; vector autoregressions; uncertainty; inflation sensitivity.
JEL Classification: C22; C24; E31.

1 Introduction

This paper aims to quantify the finite-sample behavior of state-dependent impulse response estimators under uncertainty. I begin from a design in which the true data generating process and its transition process are known, then introduce misspecifications step by step to mimic practical information constraints. The target impulse response is defined in a potential outcome framework, while estimation proceeds with prediction based, orthogonalized methods. I compare single estimators and model-averaging schemes across regimes and shock sizes, using the continuously ranked probability score (CRPS) to track how errors change as misspecifications accumulate. The goal is to provide evidence on how the transition process,

*Department of Economics, University at Albany – State University of New York, Albany, NY 12222, United States. E-mail address: sjung9@albany.edu

shock magnitude, and estimator class shape finite-sample performance in state-dependent settings.

This paper makes three contributions. First, it separates two frameworks for nonlinear impulse responses: the target impulse response is defined as a causal effect in a potential outcome setup, whereas estimation methods with orthogonalized schemes. this division serves as the organizing framework for the analysis. Second, using a controlled simulation design with stepwise misspecification across regimes and shock sizes, the paper compares VAR- and LP-based estimators alongside model-averaging schemes, offering tentative guidance for prediction-based estimation under both small and large shocks. Third, it proposes a set of novel inflation-related regimes, grounded in the literature and organized systematically, and shows that uncertainty and monetary shocks exhibit distinct, state-dependent dynamics across these regimes.

Outline The rest of the paper is as follows: Section 2 illustrates properties of state-dependent estimators and present the different frameworks to estimate impulse responses under nonlinear environment. Section 3 examines the results from Monte-Carlo simulation with uncertainty with various size of the shocks. Section 4 applies nonlinear regime into the real world with two shocks: uncertainty and monetary shocks. Section 5 is conclusion.

2 Methodology

To illustrate the properties of state-dependent estimators, I consider a structural VAR whose coefficients are conditional on the lagged state. Let

$$w_t \equiv (x_t,\ \bar{w}_t')' \in \mathbb{R}^{1+N}, \quad \bar{w}_t = (y_{1t}, \ldots, y_{Nt})',$$

where $\bar{w}_t$ collects the endogenous macroeconomic variables and the scalar x_t links the system to the policy innovation of interest, ε_{1t}: it can be the observed shock itself ($x_t = \varepsilon_{1t}$), serially correlated process or instrument. A popular way to obtain observed shocks is the narrative approach (e.g., Ramey (2011), Romer and Romer (1989)). Then, in each state, the state-dependent SVAR is

$$C_{t-1} w_t = k_{t-1} + B_{t-1}(L) w_{t-1} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ denotes mutually independent structural shocks with $\varepsilon_t \sim$ i.i.d. $(0, \Sigma)$, and Σ is diagonal with entries σ_i^2. Here, $B_{t-1}(L) = B_{1,t-1} + B_{2,t-1}L + \cdots + B_{p,t-1}L^{p-1}$ and p denotes the lag order. For simplicity, I specialize to a bivariate SVAR with $w_t = (x_t, y_t)'$ (so $\bar{w}_t = y_t$) and, in this section, I set $x_t = \varepsilon_{1t}$, the observed policy shock.

Then, coefficients evolve conditional on the S_{t-1}:

$$\begin{aligned} k_{t-1} &= (1 - S_{t-1})k_0 + S_{t-1}k_1, \\ C_{t-1} &= (1 - S_{t-1})C_0 + S_{t-1}C_1, \\ B_{j,t-1} &= (1 - S_{t-1})B_{j0} + S_{t-1}B_{j1} \quad \text{for } j = 1, \ldots, p, \end{aligned}$$

where S_t is a stationary time series. If S_t is defined by an indicator function, it takes the value 1 in favorable states (e.g., expansions) and 0 otherwise. The lag polynomial is $B_{t-1}(L)$ with coefficient matrices $\{B_{j0}, B_{j1}\}_{j=1}^{p}$. In this paper I focus on the response of y_t to a one-time innovation in ε_{1t} over time in this state-dependent SVAR.

2.1 Regime Transition Processes in Nonlinear Model

In state-dependent model, the regime is determined by a transition process, S_t, defined as a deterministic function that takes a value s. Depending on the researcher's choice, s can be binary $s \in \{0, 1\}$ or continuous $s = [0, 1]$. Let

$$\mathfrak{w}_r = (w_r', q_r)' : r \leq t \quad \text{for } r \leq t,$$

where w_r are model variables (and their lags) and q_t is an externally determined, strictly exogenous variable (e.g., measures of sentiment or *animal spirits*; see Barsky and Sims (2012), Blanchard (1993), and Hall (1993)). The transition function S_t is

$$S_t = \eta(\mathfrak{w}_r : r \leq t), \tag{2}$$

where $\eta(\cdot)$ is a researcher-specified function that maps $\{\mathfrak{w} : r \leq t\}$ to S_t. If S_t is constructed using only q_t (so $\mathfrak{w}_r = q_r$, then $S_t = \eta(q_r)$ is exogenous to the data-generating process. If instead S_t is constructed from elements of w_t (for example, $\mathfrak{w}_r = y_r$), then $S_t = \eta(y_r)$ is endogenous. A standard binary example is $S_t = \mathbf{1}(y_t > y^*)$, with threshold y^*. Employment/unemployment rates and real activity measures are common choices for transition variables when defining the state of the economy. Although some empirical studies treat S_t as exogenous, that assumption is often empirically implausible as the state is typically constructed from variables that are directly or indirectly related to the model's endogenous block.

2.2 Computing impulse responses

In nonlinear settings, impulse responses are obtained using two broad approaches. The first is a generalized, path-dependent approach that conditions on histories and allows the state to evolve along the path. It is the causal effect of a one-time shock ε_{1t} on y_{t+h}, conditional on the state of the economy at time $t-1$, S_{t-1}. The second is an orthogonalized, state-conditional approach that mirrors the linear practice (e.g., LP- and VAR-based evaluations applied by state).

2.2.1 Generalized methods

Nonlinear models are path dependent, and population impulse responses are generally not available in closed form from estimated coefficients alone. A common approach in the literature[1] defines the true response by comparing two sample paths for the outcome of interest: one in which ε_{1t} receives a one-time shock at time t, and one in which it does not.

I summarize this object as the conditional average response (CAR)[2]:

$$CAR_{y,h}(\delta, s) = E\Big[y_{t+h}(\varepsilon_{1t} + \delta) - y_{t+h}(\varepsilon_{1t})|S_{t-1} = s\Big], \tag{3}$$

where $s \in \{0, 1\}$, $y_{t+h}(\varepsilon_{1t})$ denotes the realized value along the baseline path, and $y_{t+h}(\varepsilon_{1t}+\delta)$ denotes the counterfactual that would obtain if the one-time innovation at t were shifted by δ. Equivalently, for any realization $\varepsilon_{1t} = e$, the baseline holds e fixed while the counterfactual evaluates $e + \delta$; all other shocks retain their realized values. Then, the two sample paths are:

1. Baseline path $\{y_{t+h}\}$: **observed** potential outcomes.
   $\mathcal{E} \cup \mathcal{S} = \{\ldots \varepsilon_{1t-1}, \varepsilon_t, \varepsilon_{1t+1}, \ldots, \varepsilon_{2t-1}, \varepsilon_{2t+1}, \ldots\} \cup \{\ldots S_{t-1}, S_t, S_{t+1}, \ldots\}$.

2. Alternative path $\{y^*_{t+h}\}$: **unobserved** counterfactual values. [3]
   $\mathcal{E}^* \cup \mathcal{S}^* = \{\ldots \varepsilon_{1t-1}, \varepsilon^*_t, \varepsilon_{1t+1}, \ldots, \varepsilon_{2t-1}, \varepsilon_{2t+1}, \ldots\} \cup \{\ldots S_{t-1}, S^*_t, S^*_{t+1}, \ldots\}$, where $\varepsilon^*_{1t} = \varepsilon_{1t} + \delta$ and all other shocks equal their baseline realizations.

If the states are exogenous, then $S^*_r = S_r$ for all r. If the state is endogenous, then $S^*_r \neq S_r$ for $r \geq t$ as the one-time shock at t feeds back into the state, so coefficients evolve differently along the two paths. Throughout, I condition on the lagged state S_{t-1} and do not condition on current or future states, consistent with applied macro practice and the LP literature on state-dependent responses.

The above paths are based on the assumption that the conditional impulse response of y_{t+h} depends on the state of the economy at time $t-1$, but not on the current, future or prior periods.[4]

2.2.2 Orthogonalized methods

In contrast to generalized methods, where impulse responses are defined as differences between two future outcomes under different paths, orthogonalized methods recover state-conditional impulse responses by isolating the structural shock ε_{1t} and tracing its effect through shock-identified linear-system estimators. To distinguish this object from the generalized response, I denote the orthogonalized response by $\hat{\theta}_h(s)$, where s indexes the lagged state $S_{t-1} = s$. In practice these methods typically normalize the shock scale to one standard deviation, so $\frac{\delta}{\sigma_1} = 1$ and δ is not shown explicitly; this convention is convenient

[1] Francis, Owyang, and Soques (2023); Gallant, Rossi, and Tauchen (1993); Gonçalves, Herrera, Kilian, and Pesavento (2021, 2022, 2024); Gourieroux and Jasiak (2005); Kilian and Vigfusson (2011); Koop, Pesaran, and Potter (1996); Potter (2000).

[2] Gonçalves et al. (2024)

[3] Although $\{y^*_{t+h}\}$ is unobserved, it can be recovered from the structural model given $\mathcal{E}^*$ and $\mathcal{S}^*$.

[4] This setting is more common in applied macroeconomics and the empirical practice in the LP literature on estimating state-dependent responses.

but can matter when effects are nonlinear in shock size. Two widely used implementations are local projections (LP; Jordà and Taylor (2025)) and vector autoregressions (VAR; Sims (1980)), applied by state.

State-dependent VAR In the VAR framework, impulse responses are obtained by estimating the reduced form and mapping reduced-form innovations to structural shocks via an identification scheme (e.g., Cholesky). The reduced-form representation implied by (1) is

$$w_t = C_{t-1}^{-1} k_{t-1} + \sum_{l=1}^{p} A_l w_{t-l} + u_t,$$

where $A_{l,t-1} \equiv C_{t-1}^{-1} B_{l,t-1}$, and $u_t \equiv C_{t-1}^{-1} \varepsilon_t$. Then, conditional on $S_{t-1} = s$, the moving average representation is

$$w_{t+h} = \sum_{j \geq 0} \Psi_{s,j} u_{t+h-j}, \quad \Psi_{s,0} = I,$$

where $\{\Psi_{s,j}\}$ are implied by $\{A_{l,t-1}\}_{l=1}^{p}$. The impulse response at horizon h to a one-unit change in the policy shock is the $(i,1)$-th element of $\Psi_{s,h} C_{t-1}^{-1}$.

$$\hat{\theta}_{y,h}^{\text{VAR}}(s) = \left(\Psi_{s,h} C_{t-1}^{-1}\right)_{i,1} | S_{t-1} = s, \quad i = 2, \ldots, N+1, \tag{4}$$

since y is the i-th element of w and the policy shock occupies the first position. In the bivariate case, $i = 2$. When the shock is set to $\varepsilon_{1t} = \delta$, the VAR response scale linearly, $\hat{\theta}_{y,h}^{\text{VAR}}(s;\delta) = \hat{\theta}_{y,h}^{\text{VAR}}(s;1)\delta$, and it also equals the difference between the iterated multi-step forecasts $E\big[y_{t+h} | \varepsilon_{1t} = \delta, S_{t-1=s}\big] - E\big[y_{t+h} | \varepsilon_{1t} = 0, S_{t-1=s}\big]$.

State-dependent LP In the LP framework, the impulse response is obtained by projecting the future outcome y_{t+h} on the observed policy shock at time t. In this sense, LP measures impulse responses by orthogonalized projection: the regression coefficient gives the response of the system to the shock, conditional on controls. For each horizon h, estimate

$$\begin{aligned} y_{t+h} =& (1 - S_{t-1})\big[\mu_{0,h} + \beta_{0,h}\varepsilon_{1t} + \gamma_{0,h}' \cdot controls_t\big] \\ &+ (S_{t-1})\big[\mu_{1,h} + \beta_{1,h}\varepsilon_{1t} + \gamma_{1,h}' \cdot controls_t\big] + \text{residual}_{t+h}, \end{aligned} \tag{5}$$

where coefficients are state-contingent (e.g., $\beta_h(S_{t-1}) = (1 - S_{t-1})\beta_{0,h} + S_{t-1}\beta_{1,h}$). The state-specific LP response equals the projection coefficient and can be written as

$$\hat{\theta}_{y,h}^{\text{LP}}(s) = \frac{E\big[y_{t+h}\varepsilon_{1t} | S_{t-1} = s\big]}{E\big[\varepsilon_{1t}^2 | S_{t-1} = s\big]}, \tag{6}$$

which follows from the linear projection property under the observed-shock setup. It is also equal to the difference between iterated multi-step forecasts when the shock is set to $\varepsilon_{1t} = \delta$, i.e., $E\big[y_{t+h} | \varepsilon_{1t} = \delta, S_{t-1} = s\big] - E\big[y_{t+h} | \varepsilon_{1t} = 0, S_{t-1} = s\big]$. For a shock of size δ,

when responses scale linearly, $\hat{\theta}^{\mathrm{LP}}_{y,h}(s;\delta) = \hat{\beta}_{s,h}\delta$.

State-dependent Model Averaging Suppose there are M candidate models. For model m, let $\hat{\theta}^{(m)}_{y,h}(s)$ denote the model-specific estimator of the target response at horizon h in state s. Let the weight vector be

$$\mathbf{w}_{h,s} = \big(\omega_{1,h,s}, \ldots, \omega_{M,h,s}\big)', \omega_{m,h,s} \geq 0, \sum_{m=1}^{M} \omega_{m,h,s} = 1.$$

Then, the model-averaged estimator is

$$\hat{\theta}^{mavg}_{y,h}(s) = \sum_{m=1}^{M} \omega_{m,h,s}\hat{\theta}^{m}_{y,h}(s). \tag{7}$$

Here, *mavg* denotes a generic model-averaging scheme (e.g., Bayesian model averaging, cross-validated averaging); weights may vary by horizon and state through $\mathbf{w}_{h,s}$.

3 Simulation Study

This section evaluates the finite-sample behavior of state-dependent impulse-response estimators in a setting that mirrors applied macroeconomic practice. I consider a binary regime indicator S_t that is endogenous. Also, to estimate target impulse responses, I restrict the study to orthogonalized estimators (rather than generalized estimators), their variants, and model-averaging schemes that combine them.

If, in practitioners' data, S_t is not strictly exogenous and $\delta \neq 1$, not only may state-dependent estimators fail to recover the true impulse response, but the finite-sample bias can also be sizable. (Gonçalves et al., 2024) show that, under endogenous S_t, the finite-sample bias of LP can be sizable and depends on the standardized shock. This situation may not very different for the regime-dependent VAR estimator. However, in practice, many studies report orthogonalized impulse responses scaled to a unit shock, $\delta = 1$, without considering the size of the shock relative to the standard deviation $\frac{\varepsilon_{1t}}{\sigma_1}$. To highlight the significance of this issue, I vary δ while fixing $\sigma_1 = 1$.

The simulation is designed to answer the following questions: (i) under specific forms of uncertainty, which estimator (VAR, LP, or their variants) delivers lower bias and dispersion; (ii) whether simple model averaging improves performance relative to its components; and (iii) how conclusions vary with the shock magnitude. I begin from a benchmark with complete information about the data-generating process. I then introduce realistic uncertainties sequentially, for example about the transition mechanism and lag order, and vary δ, while preserving the common empirical convention that reported impulse responses are interpreted as if $\delta = 1$. The remainder of the section details the data-generating processes, estimators, evaluation metrics and simulation results under uncertainty scenarios.

3.1 Simulation Design and DGP

In this subsection, I compare the impulse responses of model averaging and single estimators under a set of misspecifications and non-negligible shock sizes, with $\delta \in \{1, 5, 10, -1, -5, -10\}$. Motivated by Ramey and Zubairy (2018), I adopt a trivariate state-dependent SVAR(p) as the data generating process (DGP), where x_t follows an AR(p) without a constant. Here, ε_{1t} is interpreted as the structural fiscal policy shock (government spending news). Let $w_t = (x_t, g_t, y_t)'$ where x_t denotes the Ramey–Zubairy military spending news series, g_t is real government spending, and y_t is real GDP. The calibration targets quarterly U.S. data from 1890Q1 to 2015Q4.

$$
\begin{aligned}
x_t &= \sum_{j=1}^{p} \alpha_{j,t-1} x_{t-1} + \varepsilon_{1t}, \\
C_{t-1} \begin{bmatrix} x_t \\ g_t \\ y_t \end{bmatrix} &= k_{t-1} + B_{t-1}(L) \begin{bmatrix} x_{t-1} \\ g_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix},
\end{aligned} \tag{8}
$$

where ε_t are mutually independent $N(0,1)$, so $\delta = 1$ corresponds to a one standard deviation shock. Also, all lag orders are 4. Parameter values used in the simulations are reported in the online appendix.

Assume that $w_t = (x_t, g_t, y_t)'$ where x_t is Ramey and Zubairy's military spending news measure, g_t is real government spending, and y_t is real GDP in period t. All of them are quarterly data from 1890Q1 to 2015Q4.

I consider two DGPs, distinguished by the state-transition process for S_t.

1. DGP I: TVAR

   The regime is selected deterministically by a threshold rule on the transition variable z_t:

   $$
   S_t = \begin{cases} 1 & if \quad z_t > z^* \\ 0 & if \quad z_t \leq z^*, \end{cases} \tag{9}
   $$

   with the favorable regime when $z_t > z^*$ ($S_t = 1$) and the unfavorable regime otherwise ($S_t = 0$). Once z_t is given, there is no randomness in switching; regime assignment is purely threshold-driven. In the baseline specification, z_t is set equal to y_t and the threshold is fixed at $z^* = 1$.

2. DGP II: TVTP-MSVAR (Billio, Casarin, Ravazzolo, & Van Dijk, 2016; Diebold, Lee, Weinbach, et al., 1993; Filardo, 1994; Francis et al., 2023)

   The regime is selected probabilistically. A two-state Markov chain governs switching, and the transition probabilities vary over time with the observable z_{t-d}. The logistic function maps movements in z_{t-d} to the time-t transition probabilities. Thus, the same value of z_{t-d} does not necessarily imply the same next-period regime. In contrast

to the TVAR case, z_{t-d} shifts probabilities rather than imposing a deterministic cutoff. The lag is set to $d = 1$.

$$\Pr[S_t = j | S_{t-1} = i, z_{t-d}] = \pi_{ij,t} = \frac{\exp(\bar{\gamma}_{ji} + \gamma_{ji} z_{t-d})}{\exp(\bar{\gamma}_{0i} + \gamma_{0i} z_{t-d}) + \exp(\bar{\gamma}_{1i} + \gamma_{1i} z_{t-d})},$$

where $\pi_{i0,t} + \pi_{i1,t} = 1$ for each origin state $i = \{0, 1\}$. Given S_{t-1} and $\pi_{ij,t}$, assuming $u_t \sim U(0, 1)$ is the random draw. And assign:

$$S_t = \begin{cases} 1, & \text{if } S_{t-1} = 0 \text{ and } u_t < \pi_{01,t}, \\ 1, & \text{if } S_{t-1} = 1 \text{ and } u_t \geq \pi_{10,t}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

I compare estimation methods for recovering the true impulse response in Monte Carlo simulations. Starting from a benchmark with complete information where the DGP is known, I introduce misspecifications sequentially—incorrect lag order, misspecified transition mechanism, misspecified transition variable, and linear (rather than nonlinear) dynamics. Results are based on $1,000$ Monte-Carlo replications. For each replication, a long path $T^L = 1,000$ simulated to compute CAR, the first 250 observations are burned in, leaving $T = 750$ for estimation.

3.2 True Impulse Response (CAR) and Estimators

The population impulse response is defined as CAR which is obtained by generalized method introduced in Section 2.2.1. CAR and the estimation methods that recover it differ in how they identify the impulse response: In this paper, CAR is defined under the potential-outcomes framework, whereas the estimators are framed under the prediction framework in 2.2.2.

True Impulse Response The population non-linear impulse response is defined as follows:

$$\begin{aligned} CAR_{g,h}(\delta, s) &= E\big[g_{t+h}(\varepsilon_{1t} + \delta) - g_{t+h}(\varepsilon_{1t}) | S_{t-1} = s\big], \\ CAR_{y,h}(\delta, s) &= E\big[y_{t+h}(\varepsilon_{1t} + \delta) - y_{t+h}(\varepsilon_{1t}) | S_{t-1} = s\big], \end{aligned} \tag{11}$$

where $g_{t+h}(\varepsilon_{1t})$ and $y_{t+h}(\varepsilon_{1t})$ denote realized future values, while $g_{t+h}(\varepsilon_{1t}+\delta)$ and $y_{t+h}(\varepsilon_{1t}+\delta)$ denote counterfactual values that would have been observed if one-time shock ε_{1t} were perturbed by δ, conditioning on $S_{t-1} = s$.

Estimation Models. I estimate true non-linear impulse response using regression-based projections using seven single estimators (VAR- and LP-based) and transformed traditional model averaging schemes following Hounyo and Jung (2025). All estimators target the same $\text{CAR}_{\cdot,h}(\delta, s)$ for DGP defined in 3.1. Single estimators are grouped by methodological class: VAR- and LP-based. Model-averaging schemes are grouped by the class of estimators they

combine.

1. Single estimators

    - VAR-variants: Plain VAR, Bias-corrected VAR (abbreviated "BC VAR"), and Bayesian VAR (abbreviated "BVAR").
    - LP-variants: Plain LP, Bias-corrected LP (abbreviated "BC LP"), Bayesian LP (abbreviated "BLP"), and Penalized LP (abbreviated "Pen.").

2. Model Averaging (abbreviated "MAVG")

    - $\mathbf{MAVG_{VAR}}$: GMA_{VAR}, and CVA_{VAR}.
    - $\mathbf{MAVG_{LP}}$: GMA_{LP}, and CVA_{LP}.
    - $\mathbf{MAVG_{ALL}}$: $\alpha_{\text{LP,MSPE}}$-GMA, $\alpha_{\text{LP,MSPE}}$-CVA, $\alpha_{\text{LP,RS}}$-GMA, and $\alpha_{\text{LP,RS}}$-CVA.

For model averaging groups, MAVG_{VAR} aggregates VAR-based estimators, MAVG_{LP} aggregates LP-based estimators, and MAVG_{ALL} aggregates the full set. Also, within these groups, generalized model averaging (GMA) combines predictions from multiple models using weights proportional to their posterior probabilities; for GMA_{VAR} this coincides with Bayesian model averaging (BMA) in practice. By contrast, GMA_{LP} replaces the likelihood with a HAC-based criterion for weighting. The second model-averaging method, cross validation model averaging (CVA), follows a predictive approach using a leave-h-out strategy. The α_{LP} schemes in MAVG_{ALL} indirectly pool VAR- and LP-based estimators by introducing a second-stage weight α_{LP}, which allocates an overall share to the LP class; α_{LP} can be set by MSPE or R^2. This two-stage construction facilitates combining VAR and LP when their structures prevent direct first-stage pooling. [5] Additional model specifics are provided in the online appendix.

All estimation models include four lags of the data vector w_t as controls, following Ramey and Zubairy (2018). Also, nonlinear impulse responses are estimated by splitting the sample into two regimes. Consequently, except for GMA_{VAR}, and CVA_{VAR}, estimator outputs and model-averaging weights are regime- and horizon-specific. [6]

3.3 Performance Metric

I evaluate estimator performance using the continuously ranked probability score (CRPS, Francis et al. (2023); Krüger, Lerch, Thorarinsdottir, and Gneiting (2021); Matheson and Winkler (1976), which delivers a single, comparable number for each horizon, regime, and response variable. For response $i \in \{g, y\}$ at horizon h and regime s,

$$CRPS_i(F, h, s) = \int_{-\infty}^{\infty} \left[F(\mathbb{Z}) - \mathbf{1}(\mathbb{Z} > CAR_i(h, \delta, s)\right]^2 d\mathbb{Z}. \tag{12}$$

[5] For further details, see Hounyo and Jung (2025).

[6] Weights for GMA_{VAR}, and CVA_{VAR} are regime-specific but constant across horizons. All model-averaging weights are chosen by in-sample fit.

Here, $F(\mathbb{Z})$ is the posterior-predictive cumulative distribution function (CDF) of the estimator's impulse response, and $\mathbf{1}(\mathbb{Z} > CAR_i(h, \delta, s)$ is the empirical CDF that places unit mass to the realized true value $CRPS_i(F, h, s)$. CRPS measures the squared distance between the two CDFs over the real line, so smaller values indicate better performance; the score is zero only if the predictive distribution assigns all probability to the realized outcome.

For comparison across designs, I report CRPS averaged over horizons, regimes, and response variables. When aggregating across regimes, I use equal weights (0.5 for each state), yielding a single summary number per estimator. For complementary evaluation metrics, I also report absolute bias ($aBias$) and standard deviation (SD):

$$aBias(\hat{\theta}_{i,h}(s)) = \left|E(\hat{\theta}_{i,h}(s)) - CAR_i(h, \delta, s)\right|, \quad i \in \{g, y\}, \tag{13}$$

and

$$SD(\hat{\theta}_{i,h}(s)) = \left(E[\hat{\theta}_{i,h}(s) - E[\hat{\theta}_{i,h}(s)]]^2\right)^{1/2}, \quad i \in \{g, y\}, \tag{14}$$

All metrics in this paper are computed at the model (estimator) level.

3.4 Simulation Results

This subsection presents the performance of the estimators under uncertainty. I restrict attention to projection-based estimators with orthogonalized shocks, including both frequentist and Bayesian variants. Also, I adopt the common reporting convention that IRFs are scaled to a unit structural shock— implicitly treating $\frac{\delta}{\sigma_1} = 1$—and then examine how results change as δ increases.

Main Results As the main results, I primarily discuss the DGP TVAR(4) with $\delta = 1$ and report average CRPS across horizons and regimes. Result tables for the Markov-switching case and for other shock sizes are reported in the online appendix.

Table 2 reports CRPS for impulse responses obtained from the single estimators and the corresponding model-averaging schemes under a range of misspecifications when the true transition mechanism is threshold-based and the shock size is 1. Values are averaged over regimes and horizons and normalized by the CRPS from the correctly specified TVAR(4). Thus, the CRPS for Plain VAR under no misspecification equals 1.

The first column lists the misspecification type: "None" indicates correct specification (estimators may differ by researchers' preference but the DGP features are correctly imposed); "Transition process" replaces the threshold rule with Markov switching; "State variable z_t" uses a transition variable uncorrelated with the true z_t, "Lag order $p = 1$" and "Lag order $p = 6$" impose 1 or 6 lags when the true model has 4; and "Linear" estimates a linear model in place of the nonlinear DGP.

The table yields several findings for the TVAR(4) DGP under uncertainty. First, Bayesian estimators perform well across misspecifications although VAR(4) is the data-generating model. This can occur as plain VAR is an orthogonalized frequentist estimator, thus BVAR

may outperform in finite samples, with shrinkage regularizing high-dimensional parameters and stabilizing estimation under regime splits. Between BVAR and BLP, BVAR is the safer choice under misspecification, as indicated by the complementary evidence in the bias heatmaps and standard-deviation panels (Figure 1, Figure 2). In the bias maps, blue indicates smaller absolute bias and yellow indicates larger bias. BVAR attains the smallest or among the smallest biases across horizons and misspecification types, while BLP does not. Also, in the standard-deviation plots, relative rankings stabilize after several horizons, and BVAR generally exhibits the lowest dispersion, although it is delayed in some misspeficication cases. Bias-corrected estimators, on the other hand, do not perform as well as their plain or Bayesian counterparts in this nonlinear, stationary design, although they show better performance in nonstationary linear settings [7].

Second, when the true lag order is uncertain, the α_{LP} schemes with a short lag length are a safe choice. Their advantage is most pronounced when the lag order is truncated. This misspecification can occur with limited samples, especially after splitting the data by regime. This performance advantage often persists even when an additional misspecification is present.

Third, when nonlinear impulse responses are estimated with a linear model, the CRPS is nearly four times larger. This proportional increase is much larger in the TVAR ($\delta = 1$) case than under other shock magnitudes or DGPs. This suggests that, if researchers assume linearity and a unit shock, assumptions common in practice, the resulting impulse responses can differ substantially from the population values.

Other Sizes of the Shock CRPS tables for the threshold DGP with alternative shock sizes, $\frac{\delta}{\sigma_1} \in \{5, 10, -1, -5, -10\}$, are reported in the online Appendix. Comparing these outcomes with the main case ($\delta = 1$) yields several additional findings.

As the shock magnitude increases, CRPS rises markedly across all designs. This is expected as $CAR(\delta, s)$ scales with δ, while orthogonalized projections are calibrated to a unit shock and do not internalize the realized shock size during estimation. Moreover, for a given shock magnitude, total errors are slightly larger for negative shocks than for positive shocks.

Additionally, as the shock magnitude increases, Bayesian estimators begin to lose their advantage. Instead, plain and bias-corrected estimators tend to outperform, regardless of the shock's sign.

Comparing LP- and VAR-variants, LP-variants—especially Plain LP and BC LP—tend to outperform as the shock magnitude increases. Accordingly, when the shock is large and the true model is uncertain, model averaging restricted to LP-variants may suffice, reducing the need to combine VAR and LP in a single scheme.

MSVAR The online appendix reports CRPS tables for all values of $\frac{\delta}{\sigma_1}$ when the true transition process is Markov switching. The table format matches the threshold case, except

[7] Li, Plagborg-Møller, and Wolf (2024), Hounyo and Jung (2025)

that the row labeled "Transition process" now denotes switching by a threshold rather than by Markov switching. For comparability with the previous tables, all CRPS values are normalized by the CRPS from plain VAR (4) when the true transition process is threshold under correct specification.

The most noticeable pattern is that CRPS increases with shock magnitude, with a steeper rise than under the threshold DGP.

The results also indicate that adopting an alternative transition process (for example, threshold or linear) can reduce CRPS. Two mechanisms can produce this outcome. First, when the Markov-switching transition process is highly volatile or weakly predictable, the effective number of observations per regime shrinks, increasing parameter uncertainty within regimes. This raises the sampling variability of the estimated impulse responses across Monte Carlo replications, so in finite samples the estimated impulse response rarely matches the population response. Another mechanism is that simpler transition processes, such as threshold or linear, pool information across regimes. Pooling produces smoother impulse responses with lower sampling variance and sometimes lower bias by averaging over state uncertainty, while the correctly specified Markov-switching model can face finite-sample uncertainty about the prevailing state.

Taken together, these patterns support model averaging across transition processes (Francis et al., 2023), even when the true process is Markov switching but relatively complex.

Overall, LP variants, especially plain LP, outperform VAR estimators, except in a few cases when the lag order is truncated. A plausible reason is LP's design: it estimates each horizon by directly projecting the future outcome on the identified one-period structural shock. This approach is less sensitive to dynamic misspecification from complex transition processes and to lag-order truncation, whereas VARs depend on a fully specified propagation structure.

Summary The simulations indicate that CRPS is primarily determined by the true transition process and the shock magnitude. The simulation results under uncertainty are summarized as follows.

While LP-based estimators generally outperform VAR-based estimators as shocks grow, VARs remain competitive when the VAR lag order is truncated, performing well relative to LP in those cases.

Second, there is a clear inverse relationship between Bayesian estimators and the plain and bias-corrected estimators, although their relative performance does not vary systematically with shock size. When Bayesian methods perform better, the plain and bias-corrected estimators tend to perform worse, and vice versa. In addition, bias correction does not outperform the plain VAR or LP in this setting: the data are typically stationary in state-dependent environments, particularly when S_t is endogenous. In such cases, differencing the data can reduce the predictive power of the lagged controls, thus causing potentially large efficiency losses——even in large samples.[8]

[8] Olea, Plagborg-Møller, Qian, and Wolf (2025)

Third, I find evidence that, using a transition process that differs from the true one can reduce estimation error when regime identification is noisy (e.g., under Markov switching with high volatility or short samples) despite misspecification.

Lastly, when nonlinear models are mistakenly treated as linear, the estimated impulse responses are distorted more severely than under other forms of misspecification.

Recommendations Based on these simulations, I offer tentative recommendations for estimating nonlinear impulse responses within prediction-based frameworks across small and large shocks.

When the transition process is uncertain, adopting a simple specification (for example, a threshold) is a conservative choice that tends to reduce overall bias. For large shocks, and when researchers do not employ potential outcome based estimators such as generalized impulse responses (Koop et al., 1996), LP estimators are often preferable to VARs.

Although there is no uniform ranking between plain/bias-corrected and Bayesian estimators, their performance tends to move in opposite directions. Given this, when the dominant class is unclear ex ante, model averaging is advisable. Also, for large shocks, restricting the averaging set to LP-based estimators usually suffices. In stationary environments, bias correction is seldom warranted, so the averaging set can exclude bias-corrected variants with little risk. However, When misspecification is well characterized, a single well-chosen estimator is often more efficient.

4 Empirical Application

In this section, I use actual data to study first and second moment shocks across nonlinear regimes. In the linear regime, the effect of an uncertainty shock (a second moment shock) fades within about six months, while a monetary policy shock (a first moment shock) persists for roughly two to three years, as shown by Bloom (2009). Following this benchmark, I assign these two shocks to separate regimes rather than a single linear environment and examine how their effects evolve and transmit to real economic activity across regimes.

Inflation-related regimes Guided by the idea that "not all inflation is equal—the nature of inflation matters"[9], I define three inflation-related regimes and examine whether the transmission of shocks to real economic activity differs across these regimes.

Data Table 1 lists the variables used in the estimation models, excluding the transition variable that enters the transition function. All endogenous series from log(S&P 500 Index) through the federal funds rate are HP-detrended with $\lambda = 129,600$. The sample period differs by regime due to data availability. For the first and third inflation-related regimes, I use the series in the table from July 1962 through June 2008. For the second inflation-related regime, the sample runs from Jan 1978 through June 2008 when the transition variable is

[9] Cieslak and Pflueger (2023)

the household survey, and from July 1981 through June 2008 when transition variable is the professionals' survey. [10]

Table 1: Description of Data (Monthly)

Variable	**Source / Notes**
Actual volatility series	Bloom (2009); external instrument
log(S&P 500 Index)	S&P Dow Jones Indices (via FRED)
log(Average Hourly Earnings)	BLS (CES), production workers, SA
log(CPI-U)	BLS, all urban consumers, SA
Average weekly hours (manufacturing)	BLS (CES), SA
Employment (manufacturing)	BLS (CES), SA
log(Industrial Production, manufacturing)	FRB (G.17), SA
Federal funds rate	FRB, effective rate

Note: All series are monthly. See Bloom (2009), Appendix A, for details. The actual volatility series is constructed by splicing two measures: before 1986, it is the monthly standard deviation of daily S&P 500 returns, normalized to match the mean and variance of the VXO index during the overlap period; from 1986 onward, it is the Chicago Board Options Exchange VXO index of implied volatility (percent) on a hypothetical at-the-money S&P 100 option with 30 days to expiration.

Models I consider four models, two VAR-based and two LP-based, to estimate the effect of an uncertainty shock. I then compare their impulse-response trajectories. All responses are scaled to a one-unit volatility shock by normalizing with the impulse response of actual volatility at $h = 0$.

1. **Internal instrument VAR(12)** I estimate a VAR model using equation (1), setting the actual volatility series as the instrument. The observed variables in w_t are ordered as in Table 1.

2. **Internal instrument BVAR(12)** I retain the same ordering of w_t but estimate the reduced-form coefficients using a Bayesian approach. Coefficients are posterior means under a Minnesota-style prior, with variance hyperparameters set following Canova (2011).

3. **IV-LP(2) with contemporaneous controls** I estimate an IV LP model using (5). The controls include contemporaneous values of the endogenous variables and two lags of both the instrument and the endogenous variables.

4. **IV-LP(2) without contemporaneous controls** This specification is identical to the previous IV-LP model but excludes contemporaneous endogenous variables from the control set; only lagged controls are included.

[10] Although I use the realized volatility series as an external instrument rather than the observed shock itself, the implied δ/σ_1 average 3.08 over the sample and ranges from 1.5 to 7.82.

5. **Model averaging schemes** For comparison, I also report three combination estimators: GMA_{VAR} (averaging VAR(12) and BVAR(12)), GMA_{LP} (averaging IV-LP(2) with and without contemporaneous controls), and $\alpha_{\text{LP,MSPE}}$–GMA, which averages across all four base estimators using MSPE-based weights.

On the other hand, the monetary policy shock is identified as the last innovation from a Cholesky decomposition of the reduced-form residuals. Specifically, I employ IV/proxy identification for the second moment shock (the uncertainty shock), while I use recursive identification for the first moment shock (the monetary policy shock). Although the two shocks are identified differently, I compare their trajectories across regimes, keeping the main focus on the uncertainty shock.[11] Thus, the purpose is to compare the effects of the two shocks, not to equate their impulse responses.

4.1 Good vs. Bad inflation

Bonelli, Palazzo, and Yamarthy (2025) distinguishes between good and bad inflation, implying that a given shock can have opposite effects depending on how inflation relates to growth. For example, an uncertainty shock or policy news may expand the economy if investors view higher inflation as a signal of strong demand; however, it may contract the economy if inflation is perceived as eroding purchasing power without supporting growth. In general, good inflation is often demand-driven and associated with robust real activity: inflation is positively correlated with real variables. Bad inflation, by contrast, is typically supply-driven or associated with weak growth: it can stifle real activity and raise the risk of stagflation.

Transition process and data Boons, Duarte, De Roon, and Szymanowska (2020) show that the covariance between inflation and future consumption growth helps determine the equity-implied inflation risk premium. Motivated by this, I classify "good" and "bad" inflation regimes using the rolling correlation between inflation growth ($\Delta \ln P$) and demand growth ($\Delta \ln C$) computed over a 60-month trailing window:

$$\rho_t = \text{corr}\left(\{\Delta \ln P_{t-L+1}, \ldots, \Delta \ln P_t\}, \{\Delta \ln C_{t-L+1}, \ldots, \Delta \ln C_t\}\right), \quad L = 60. \tag{15}$$

Here, P denotes the Personal Consumption Expenditures Price Index (PCEPI)[12], and C denotes real personal consumption expenditures on nondurable goods and services (RNDURID and RSERVID)[13]. To compute the rolling window, the inputs P, RNDURID, and

[11] Normalizations differ by shock. For the uncertainty shock, I compute impulse responses to a one-unit volatility shock by dividing by the impulse response of actual volatility at $h = 0$. For the monetary policy shock, I compute impulse responses to a 1 percentage-point innovation in the federal funds rate. Thus, the normalization variable is the actual volatility series for the uncertainty shock and the federal funds rate for the monetary policy shock.

[12] Personal Consumption Expenditures: Chain-type Price Index, monthly, seasonally adjusted, Federal Reserve Bank of St. Louis (FRED). As PCEPI is available only from January 1959, I back-splice CPI monthly inflation to extend the series for the rolling window via backward recursion: $\ln(\widetilde{\text{PCEPI}}_{t-1}) = \ln(\widetilde{\text{PCEPI}}_t) - \pi_t^{\text{CPI}}$, where $\pi_t^{\text{CPI}} = \Delta \ln \text{CPI}_t$.

[13] Real Personal Consumption Expenditures: Nondurable Goods (chain-type quantity index, 2017=100) and Real Personal Consumption Expenditures: Services (chain-type quantity index, 2017=100), quarterly, converted to monthly, seasonally adjusted, Federal Reserve Bank of St. Louis (FRED).

RSERVID span July 1957 through June 2008, whereas the other series used for estimating impulse responses in Table 1 span July 1962 through June 2008.

If $\rho_t > 0$, periods of higher demand tend to coincide with higher inflation (and lower demand with lower inflation), which is consistent with demand-driven inflationary pressure; I classify such months as "good" inflation ($s = 1$). If $\rho_t \leq 0$, inflation tended to be high when demand was weak (and low when demand was strong), which I classify as "bad" inflation ($s = 0$):

$$S_t^{GB} = \mathbf{1}(\rho_t > 0) \tag{16}$$

Result Figure 3 reports impulse responses to the uncertainty shock (top panel) and the monetary policy shock (bottom panel). Shaded regions show the 68% bands from $\alpha_{\text{LP,MSPE}}$-GMA.

First, for the monetary policy shock, the regime split is weak across the inflation-related regimes (see Figures 4 and 5). The bands in $s = 0$ and $s = 1$ straddle zero for most horizons, indicating that the responses are statistically insignificant and statistically indistinguishable across states. In short, the monetary policy shock propagates in a manner close to a linear benchmark, so state dependence appears limited. Thus, I focus mainly on the uncertainty shock here and in the other regime analyses.

For the uncertainty shock, industrial production (IP) in the bad regime shows a small initial dip for $h \leq 6$, then rise above zero. In the good regime, responses decline slowly at first and begin to rebound around at $h = 11$. For the VAR-based estimators and $\alpha_{\text{LP,MSPE}}$-GMA, the negative effects begin to fade from $h = 24$ onward. Employment responses mirror IP. Overall, in the good regime the effects build more slowly and persist longer.

In both regimes, VAR traces a relatively smooth small hump and BVAR produces the largest, rounded hump. LP estimators oscillate more, reflecting higher sampling variability. This likely reflects the methods' construction: LP uses direct, horizon-by-horizon projections, whereas VAR extrapolates from a joint dynamic system. Also, although $\alpha_{\text{LP,MSPE}}$-GMA combines both VAR- and LP-based estimators, it assigns high weight to VAR-based estimators, so its path closely matches GMA_{VAR}. This likely reflects sample-length sensitivity: with short samples and fixed lag orders, LPs estimators have higher variance and more oscillatory responses than correctly specified VARs, which are more efficient in finite samples.

Robustness: GIRF Figure (will be linked soon) in Online Appendix reports results when I replace plain VAR(12) with generalized impulse response function (GIRFs) following Koop et al. (1996). In my implementation, the GIRF conditions on a one standard deviation shock calibrated to the sample standard deviation of the identified innovation, so its responses are mechanically larger than those from the BVAR, with the difference most pronounced in the bad regime.[14] Aside from level differences, the patterns closely mirror the plain VAR in terms of timing and shape.

[14] GIRF(p) uses the same model-averaging weights as plain VAR(p).

4.2 Anchored vs. Unanchored expectations

In this regime, I focus on inflation expectations and their dispersion instead of realized inflation. When expectations are well anchored and agents largely agree on the outlook, the economy may respond to shocks differently than when expectations are unanchored or widely dispersed. For example, if a shock arrives when expectations are aligned, policymakers and private agents are more likely to react in predictable ways; if the same shock arrives amid high disagreement, households and firms do not respond uniformly. Consistent with this view, Dong, Liu, Wang, and Wei (2024) find that the effects of monetary policy shocks are attenuated when inflation expectations are unanchored and dispersed.

Transition process and data To construct the transition indicator for this regime, I use the interquartile range (IQR) of inflation expectations to split states:

$$IQR_t = Q_{75,t} - Q_{25,t},$$

where $Q_{75,t}$ and $Q_{25,t}$ are the 75th and 25th percentile forecasts in month t. The IQR spans the middle 50% of the distribution and summarizes the distance between a typical optimist and a typical pessimist, ignoring the tails. The indicator is

$$S_t^{Exp} = \mathbf{1}\{IQR_t \leq q_{0.4}\}, \tag{17}$$

where $q_{0.4}$ is the 40th percentile of IQR_t^{HH} over the available sample. Months with $IQR_t^{\text{HH}} \leq q_{0.4}^{\text{HH}}$ are classified as anchored expectations ($s = 1$); otherwise, expectations are unanchored or highly dispersed ($s = 0$).

For the main results, I use the Michigan Survey of Consumers household responses (PX1_R50) to the question: "During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?" Results based on professional forecasters are reported in the Appendix.[15]

Result Figure 4 reports impulse responses to the uncertainty shock (top panel). Shaded regions show 68% bands from $\alpha_{\text{LP,MSPE}}$–GMA.

In the unanchored regime, IP declines on impact and begins to recover around $h = 6$–10. Employment mirrors IP, with state differences more statistically distinguishable. In the anchored regime, responses rise slowly and become statistically positive around $h = 20$–27. Employment shows a similar pattern with smaller magnitude.

Across both regimes, estimator features are consistent: VAR traces a smooth, small hump; BVAR yields the largest, rounded hump; LP estimators are more oscillatory. $\alpha_{\text{LP,MSPE}}$–GMA places most of its weight on VAR-based estimators, so its path closely tracks GMA_{VAR}.

[15] Survey of Professional Forecasters (SPF): Measures of cross-sectional dispersion for quarterly forecasts of the Consumer Price Index, next 12 months (quarterly; professionals; starting 1981:Q3).

Robustness: Households vs. Professionals Figure in Appendix reports results for both households and professionals over the shorter common sample. Comparing specifications that use each group's survey as the transition variable, the $\alpha_{\text{LP,MSPE}}$–GMA bands are narrower when the transition variable is based on professionals' expectations. This likely reflects lower measurement noise and more stable dispersion in professionals' expectations, which yields fewer regime misclassifications and a more balanced within-state sample—tightening standard errors.

When professionals' agreement is low, the uncertainty shock response of IP is positive, which contrasts with the negative response under low household agreement. By contrast, when agreement is high, both surveys yield positive responses, although the professional-based estimates are statistically less reliable at some horizons.

Robustness: GIRF Figure in Appendix reports GIRF results. GIRFs display larger amplitudes than the other estimators, and the uncertainty shock responses for IP and employment become statistically reliable within the $\alpha_{\text{LP,MSPE}}$–GMA bands in both regimes. In the unfavorable regime, the response is negative and near-immediate; in the favorable regime, the response is positive with a delayed onset.

4.3 Low vs. High Inflation Volatility

In this regime, I study the volatility (uncertainty) of realized inflation. A long line of research documents that higher inflation is associated with greater inflation uncertainty, and that elevated uncertainty can imped growth (Ball, 1992; Friedman, 1977). When price volatility is low, price changes are more stable and predictable. When inflation departs markedly from target, whether during deflationary episodes or high-inflation periods, greater uncertainty can discourage spending and investment by households and firms.

Transition process and data I classify inflation volatility using a symmetric threshold around the Federal Reserve's 2% target for PCEPI inflation. Let $\pi_t = 1200 \times \log(P_t/P_{t-1})$ denote the annualized monthly inflation rate. I define

$$S_t^{LH} = \mathbf{1}\big(|\pi_t - \pi^*| \leq 1\big), \qquad \pi^* = 2, \tag{18}$$

so that $S_t^{LH} = 1$ indicates low volatility ($s = 1$, inflation within a ± 1 percentage point band around target, i.e., 1-3%), and $S_t^{LH} = 0$ indicates high volatility ($s = 0$, outside the band). The sample for this regime is July 1962 through June 2008.

Result Figure 5 reports impulse responses to the uncertainty shock (top panel). Shaded regions show 68% bands from $\alpha_{\text{LP,MSPE}}$–GMA.

In the high–volatility regime, IP and employment decline on impact; the employment response is more statistically distinguishable across horizons. In the low–volatility regime, responses rise gradually and become statistically positive around $h = 20$–27.

Estimator features mirror those observed in the other inflation-related regimes.

Robustness: GIRF Figure in Appendix reports GIRF results. In the high-volatility regime, responses are negative for $h < 10$, then rebound and become significantly positive around $h = 15$–27 for GIRF, GMA_{VAR}, and $\alpha_{\text{LP,MSPE}}$–GMA; LP-based estimators, including GMA_{LP}, show the same pattern with smaller magnitudes. In the low-volatility regime, the uncertainty shock responses of IP and employment are statistically insignificant across horizons.

4.4 Summary

Examining the results, I summarize the common patterns across the three inflation-related regimes.

First, the uncertainty shock exhibits clear state dependence, whereas the monetary policy shock shows little to no state dependence and appears more consistent with a linear regime.

Second, in the favorable state ($s = 1$), responses to the uncertainty shock build more slowly than in the unfavorable state ($s = 0$). This pattern This is consistent with greater macroeconomic stability reducing immediate amplification and yielding slower adjustment.

Third, GIRFs, estimated under the potential-outcome framework, tend to be more sensitive in unfavorable states. Specifically, when plain VAR is replaced with GIRF, the uncertainty shock responses in unfavorable states become larger and are statistically significant for both the GIRF and the model averaging estimators.

Forth, $\alpha_{\text{LP,MSPE}}$-GMA estimator tends to place higher weight on VAR-based estimators across all inflation-related regimes. This likely reflects the effective reduction in sample size from splitting the data into two states, which raises the finite-sample variance of LP-based estimates; as a result, the second weight $\alpha_{\text{LP,MSPE}}$ favors the smoother VAR/BVAR paths.

It might stem from the short sample size by dividing the data into two regimes to estimate the state-dependent impulse responses, which causes high volatility of LP-based estimators. This situation would be common when researchers split samples by state to estimate impulse responses using orthogonalized methods.

5 Conclusions

This paper studies the finite-sample behavior of state-dependent impulse response estimators under uncertainty. I begin from a setting in which the true data generating process and its transition rule are known in the design, and then introduce misspecifications one at a time to mimic the information constraints faced by practitioners. Using a potential outcome framework to define the true impulse responses, I evaluate how predictive estimators—conventional in applied work—track this causal target when the researcher does not know the truth. Performance is measured by CRPS, which allows me to trace how errors expand as misspecifications accumulate and to compare model-averaging estimators that pool LP- and VAR-based variants as well as single estimators.

The paper makes three contributions. First, the paper draws a line between two frameworks for nonlinear impulse responses. The causal-effect object defines the true target

(CAR) within a potential outcome setup, while estimation proceeds with orthogonalized methods such as LP and VAR. This distinction serves as an organizing lens for the analysis. The causal-effect object defines the true target (CAR) within a potential outcome setup, while estimation proceeds with orthogonalized methods such as VAR and LP. This distinction serves as an organizing perspective for the analysis. Second, using a controlled simulation design with incremental misspecification across regimes and shock sizes, I compare VAR- and LP-based estimators together with model averaging schemes. Based on these results, I offer tentative recommendations for estimating nonlinear impulse responses within prediction based frameworks for both small and large shocks. Third, I propose a set of novel inflation-related regimes, grounded in the literature and organized systematically. In the data, uncertainty shocks and monetary policy shocks appear to generate distinct, state dependent dynamics across these regimes.

The simulation results suggest three broad messages. First, as shock size increases, LP-based estimators tend to outperform, although truncated-lag VARs remain competitive in several designs. Second, Bayesian and plain or bias-corrected estimators tend to move in opposite directions, and there is no simple pattern by regime or shock size that consistently favors one class. Third, a simple transition process can lower errors when the true transition process is noisy. Fourth, imposing linearity on a nonlinear process can produce the large distortions.

In the empirical study, I find that uncertainty shocks are state dependent, whereas monetary policy shocks show little or no state dependence and are broadly consistent with linear dynamics. In favorable states, responses to uncertainty shocks build more slowly than in unfavorable states. Replacing plain VAR with GIRFs, estimated under the potential outcome framework, makes the uncertainty-shock responses in unfavorable states larger and statistically significant for both GIRF and the model-averaging estimators.

Overall, this paper offers an empirical benchmark rather than a prescription. Researchers should proceed cautiously under uncertainty as estimator performance depends on the transition process and the shock size.

References

Ball, L. (1992). Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, *29*(3), 371–388.

Barsky, R. B., & Sims, E. R. (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review*, *102*(4), 1343–1377.

Billio, M., Casarin, R., Ravazzolo, F., & Van Dijk, H. K. (2016). Interconnections between eurozone and us booms and busts using a bayesian panel markov-switching var model. *Journal of Applied Econometrics*, *31*(7), 1352–1370.

Blanchard, O. (1993). Consumption and the recession of 1990-1991. *The American Economic Review*, *83*(2), 270–274.

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica*, *77*(3), 623–685.

Bonelli, D., Palazzo, B., & Yamarthy, R. (2025). "good" inflation,"bad" inflation: Implications for risky asset prices. *Inflation,"Bad" Inflation: Implications for Risky Asset Prices (January, 2025). FEDS Working Paper*(2025-2).

Boons, M., Duarte, F., De Roon, F., & Szymanowska, M. (2020). Time-varying inflation risk and stock returns. *Journal of Financial Economics*, *136*(2), 444–470.

Canova, F. (2011). *Methods for applied macroeconomic research.* Princeton university press.

Cieslak, A., & Pflueger, C. (2023). Inflation and asset returns. *Annual Review of Financial Economics*, *15*(1), 433–448.

Diebold, F. X., Lee, J.-H., Weinbach, G. C., et al. (1993). Regime switching with time-varying transition probabilities.

Dong, D., Liu, Z., Wang, P., & Wei, M. (2024). Inflation disagreement weakens the power of monetary policy.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, *12*(3), 299–308.

Francis, N., Owyang, M., & Soques, D. F. (2023). *Impulse response functions for self-exciting nonlinear models* (Tech. Rep.). National Bureau of Economic Research.

Friedman, M. (1977). Nobel lecture: inflation and unemployment. *Journal of political economy*, *85*(3), 451–472.

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica: Journal of the Econometric Society*, 871–907.

Gonçalves, S., Herrera, A. M., Kilian, L., & Pesavento, E. (2021). Impulse response analysis for structural dynamic models with nonlinear regressors. *Journal of Econometrics*, *225*(1), 107–130.

Gonçalves, S., Herrera, A. M., Kilian, L., & Pesavento, E. (2022). When do state-dependent local projections work?

Gonçalves, S., Herrera, A. M., Kilian, L., & Pesavento, E. (2024). State-dependent local projections. *Journal of Econometrics*, *244*(2), 105702.

Gourieroux, C., & Jasiak, J. (2005). Nonlinear innovations and impulse responses with application to var sensitivity. *Annales d'Economie et de Statistique*, 1–31.

Granger, C. W. (1995). Modelling nonlinear relationships between extended-memory variables. *Econometrica: Journal of the Econometric Society*, 265–279.

Hall, R. E. (1993). Macro theory and the recession of 1990-1991. *The American Economic Review*, *83*(2), 275–279.

Hounyo, U., & Jung, S. (2025, June). *Two-stage model averaging for impulse responses: Local projections- and vars-based approaches.* (Working paper)

Jordà, Ò., & Taylor, A. M. (2025). Local projections. *Journal of Economic Literature*, *63*(1), 59–110.

Kilian, L., & Vigfusson, R. J. (2011). Are the responses of the us economy asymmetric in energy price increases and decreases? *Quantitative Economics*, *2*(3), 419–453.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, *74*(1), 119–147.

Krüger, F., Lerch, S., Thorarinsdottir, T., & Gneiting, T. (2021). Predictive inference based on markov chain monte carlo output. *International Statistical Review*, *89*(2), 274–301.

Li, D., Plagborg-Møller, M., & Wolf, C. K. (2024). Local projections vs. vars: Lessons from thousands of dgps. *Journal of Econometrics*, *244*(2), 105722.

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management science*, *22*(10), 1087–1096.

Olea, J. L. M., Plagborg-Møller, M., Qian, E., & Wolf, C. K. (2025). *Local projections or vars? a primer for macroeconomists* (Tech. Rep.). National Bureau of Economic Research.

Owyang, M. T., Ramey, V. A., & Zubairy, S. (2013). Are government spending multipliers greater during periods of slack? evidence from twentieth-century historical data. *American Economic Review*, *103*(3), 129–134.

Potter, S. M. (2000). Nonlinear impulse response functions. *Journal of Economic Dynamics and Control*, *24*(10), 1425–1446.

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The quarterly journal of economics*, *126*(1), 1–50.

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of political economy*, *126*(2), 850–901.

Romer, C. D., & Romer, D. H. (1989). Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual*, *4*, 121–170.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Table 2: CRPS across simulation specifications, $\frac{\delta}{\sigma_1} = 1$

Misspecification	**True transition process: Threshold**														
	VARs			**LPs**				**MAVG$_{\mathbf{VAR}}$**		**MAVG$_{\mathbf{LP}}$**		$\boldsymbol{\alpha}_{\text{LP,MSPE}}$		$\boldsymbol{\alpha}_{\text{LP,RS}}$	
	Plain	BC	Bayes	Plain	BC	Bayes	Pen.	GMA	CVA	GMA	CVA	GMA	CVA	GMA	CVA
None	1.00	0.98	0.86	1.18	1.18	0.96	1.21	0.93	0.98	1.11	1.07	0.95	0.98	0.95	0.96
Transition process	2.18	2.21	2.12	2.20	2.24	1.98	2.23	2.16	2.21	2.15	2.07	2.17	2.18	2.16	2.17
State variable z_t	1.47	1.47	1.39	1.58	1.59	1.43	1.60	1.44	1.47	1.54	1.51	1.44	1.46	1.44	1.45
Lag order $p = 1$	0.99	1.01	1.03	1.21	1.21	1.18	1.22	1.01	1.01	1.19	1.20	0.97	0.97	0.95	0.94
Lag order $p = 6$	1.05	1.04	0.89	1.18	1.19	0.93	1.21	0.98	1.04	1.10	1.07	0.99	1.03	0.99	1.01
Transition process and state variable z_t	2.57	2.60	2.61	2.64	2.67	2.47	2.64	2.59	2.60	2.60	2.56	2.59	2.58	2.58	2.57
Transition process and lag order $p = 1$	2.08	2.12	2.12	2.13	2.17	2.14	2.15	2.10	2.12	2.14	2.15	2.08	2.08	2.12	2.13
Transition process and lag order $p = 6$	2.19	2.22	2.12	2.20	2.23	1.93	2.22	2.17	2.21	2.14	2.06	2.16	2.18	2.16	2.16
State variable z_t and lag order $p = 1$	1.55	1.58	1.59	1.65	1.67	1.69	1.67	1.57	1.58	1.66	1.69	1.50	1.49	1.52	1.51
State variable z_t and lag order $p = 6$	1.45	1.46	1.36	1.58	1.59	1.38	1.60	1.42	1.45	1.53	1.50	1.42	1.44	1.42	1.43
Linear	4.22	4.27	3.37	4.01	4.06	4.00	4.03	4.03	4.26	4.00	3.94	4.06	4.25	4.05	4.20

This table shows the mean CRPS for each estimation method when the true model's transition process is Threshold. I compute the overall CRPS as the average over all horizons, dependent variables, and regimes. I normalize values by setting the CRPS when the true model is estimated by plain VAR(4) without any misspecification is one. The true model is TVAR(4). An incorrect transition occurs with the Markov-switching process. The incorrect state variable is uncorrelated with the true z_t. Linear displays CRPS when I compute the true non-linear model with LSVAR(4).

TVAR: Bias of selected estimators under misspecifications, $\delta/\sigma_1 = 1$

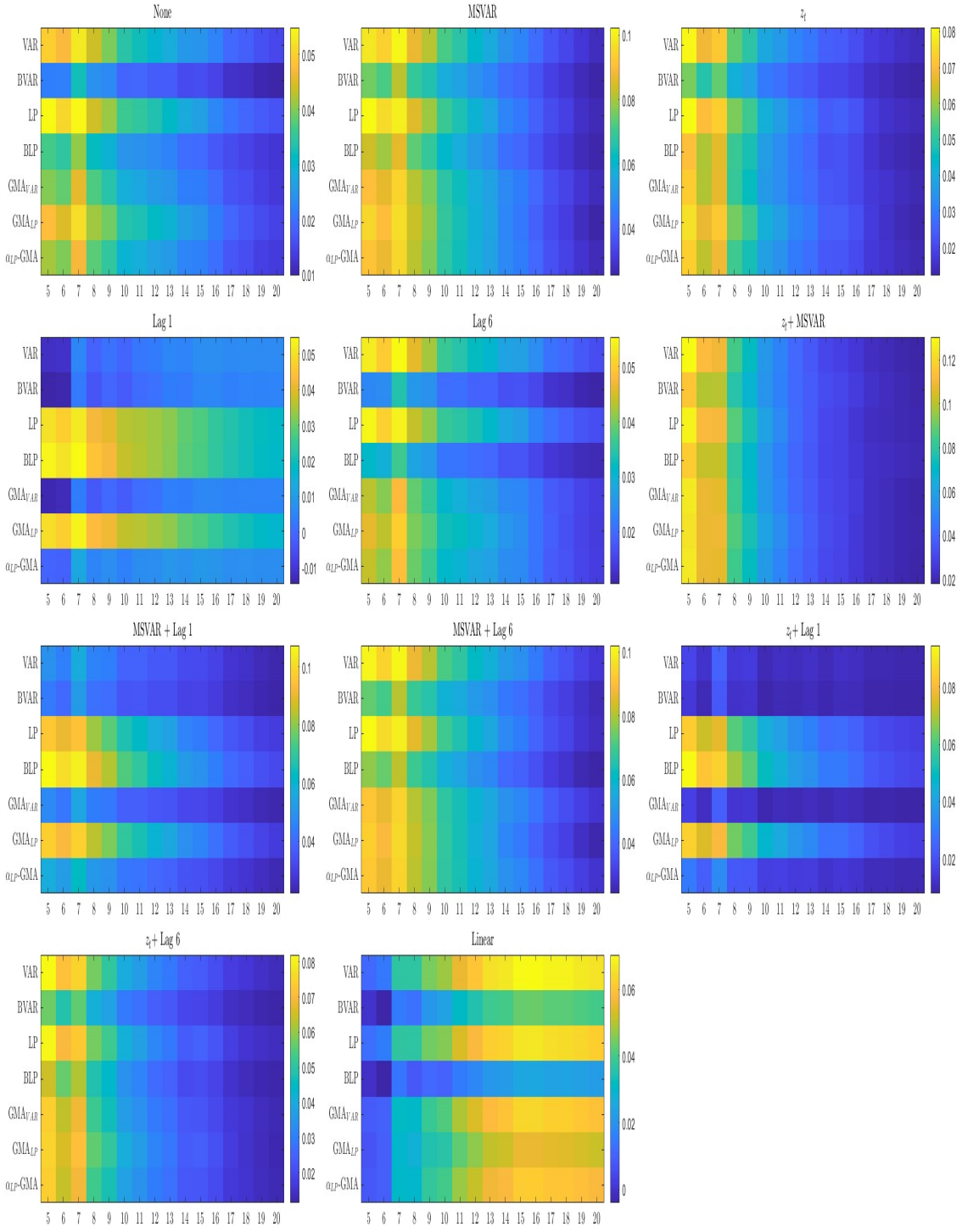


Figure 1

TVAR: Standard deviation of selected estimators under misspecifications

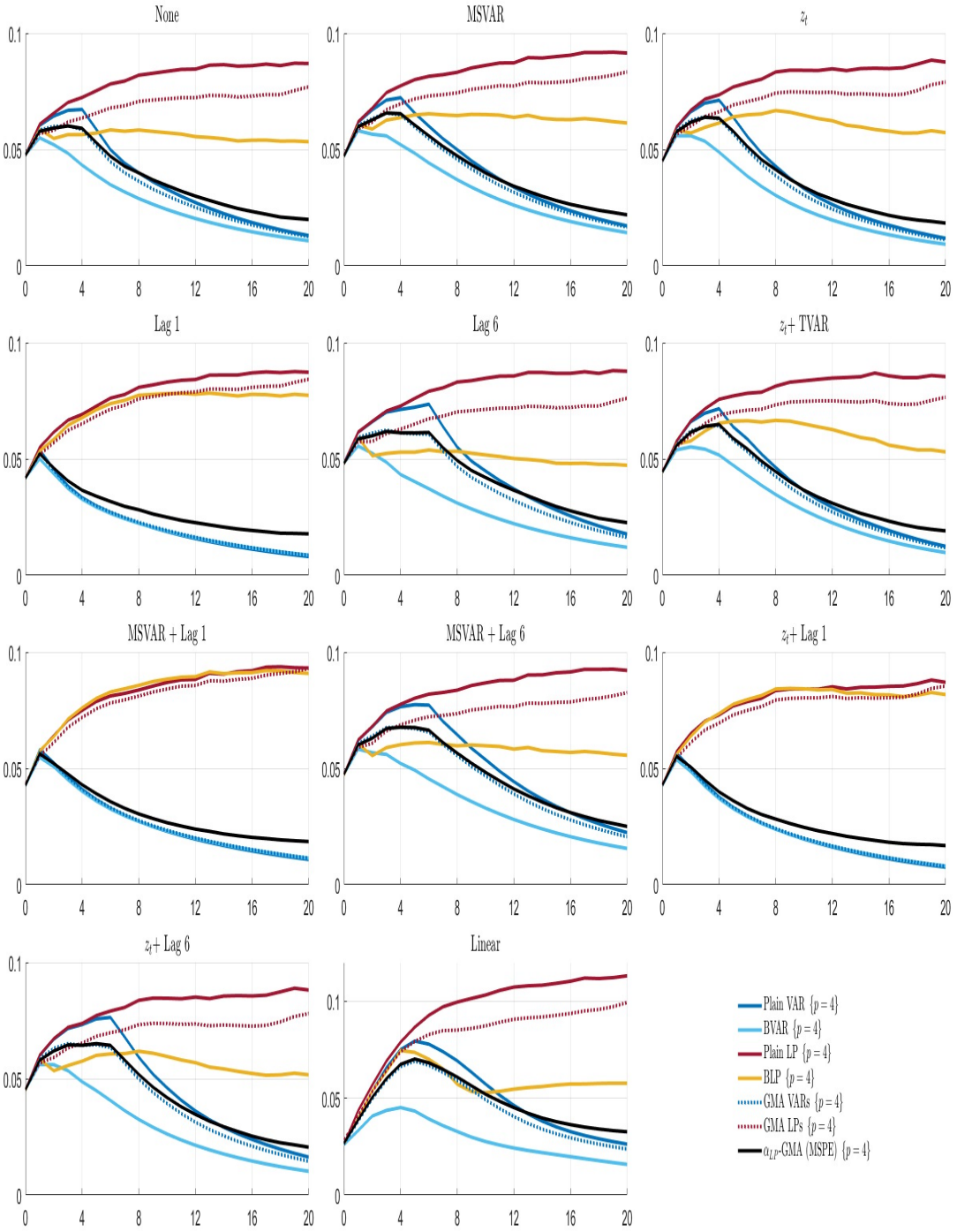
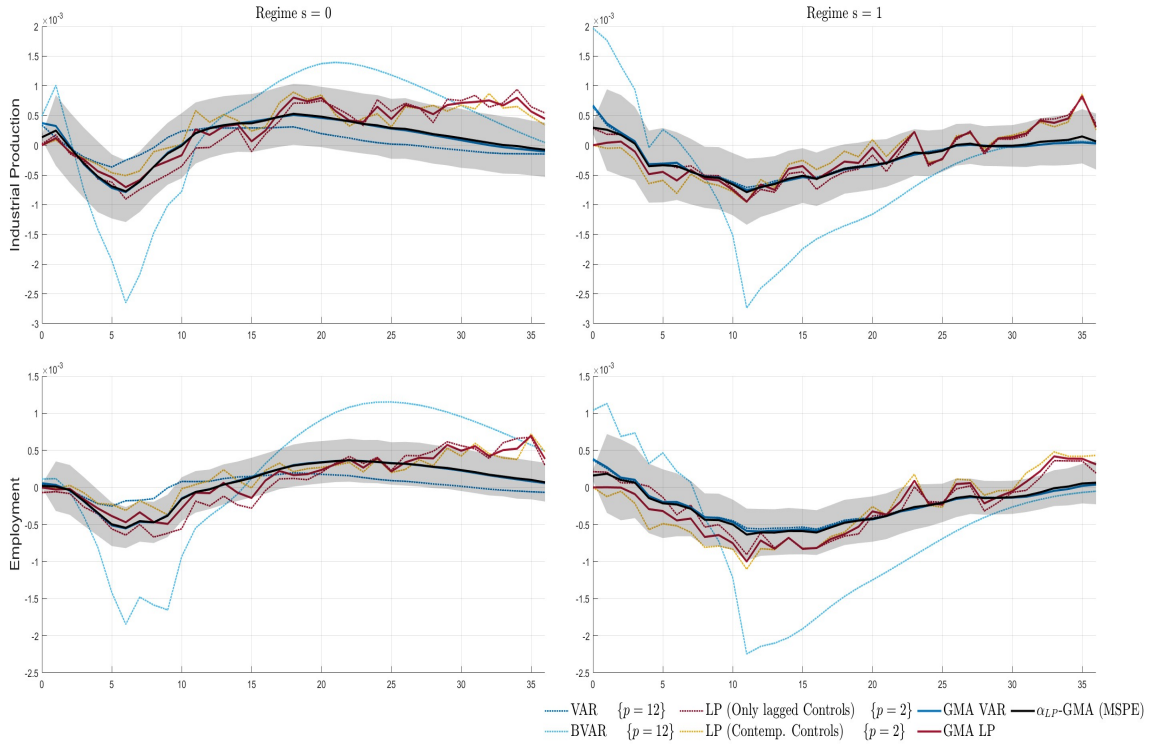


Figure 2

Good vs. Bad inflation regimes:

The impact of the uncertainty shock



The impact of the monetary shock

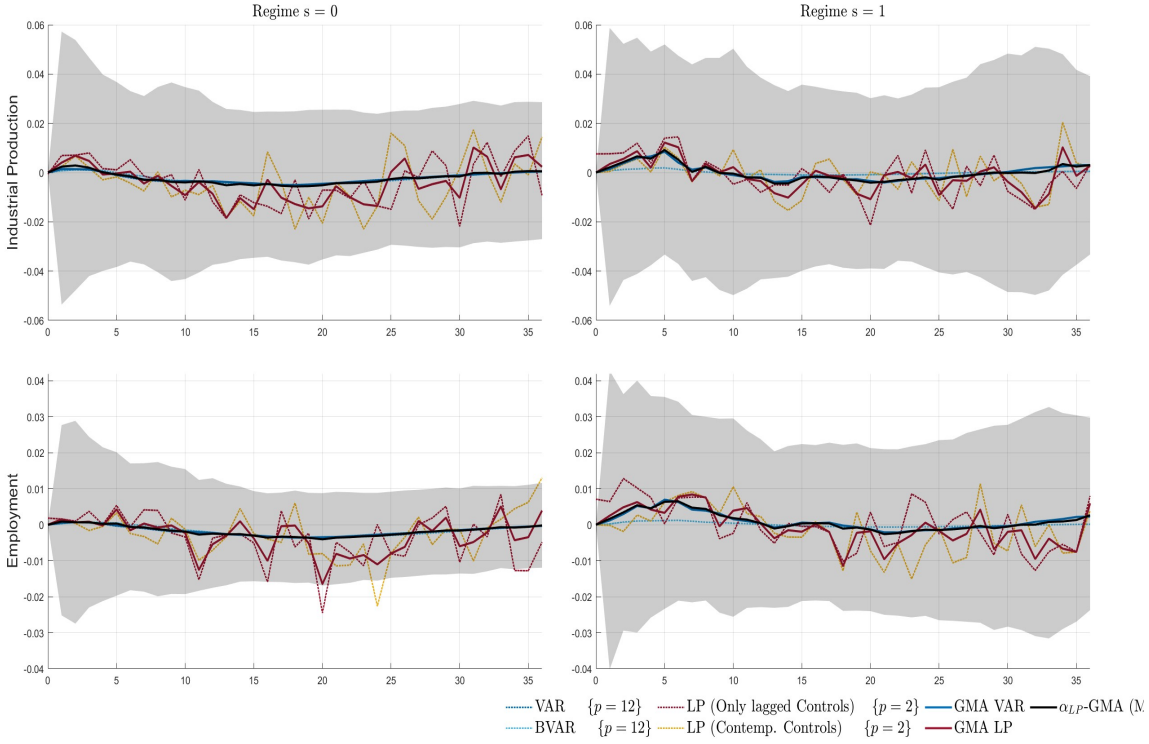
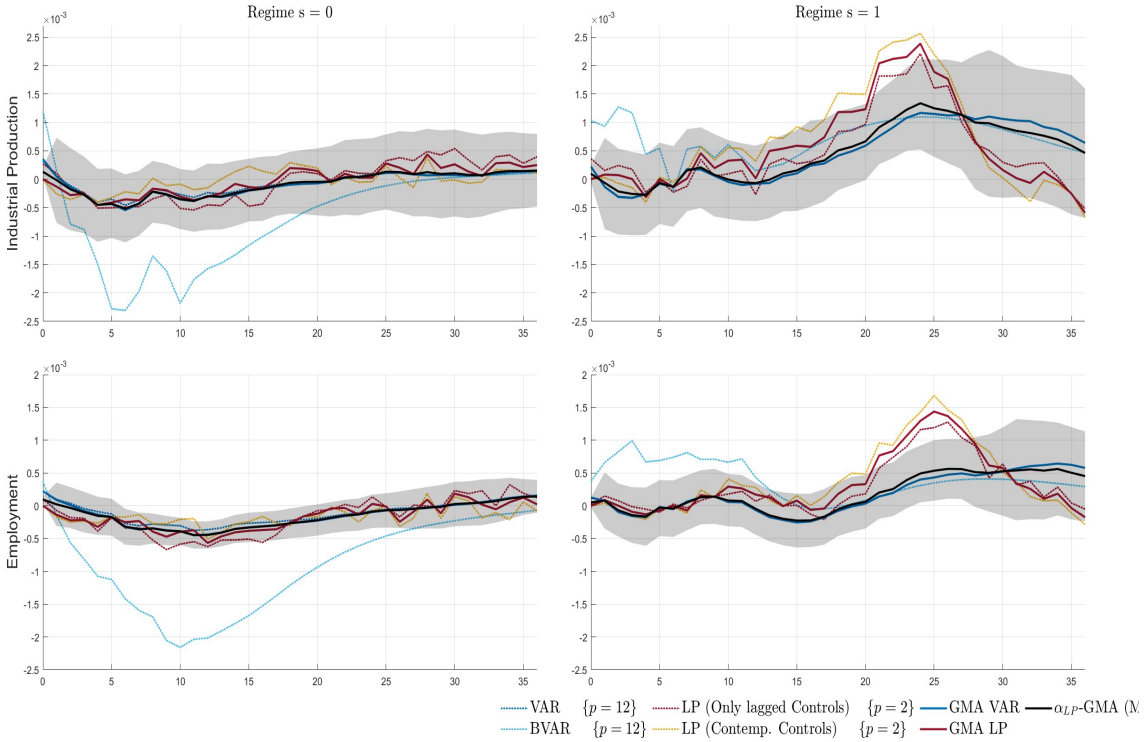


Figure 3

Anchored vs. Unanchored expectation regimes:

The impact of the uncertainty shock



The impact of the monetary shock

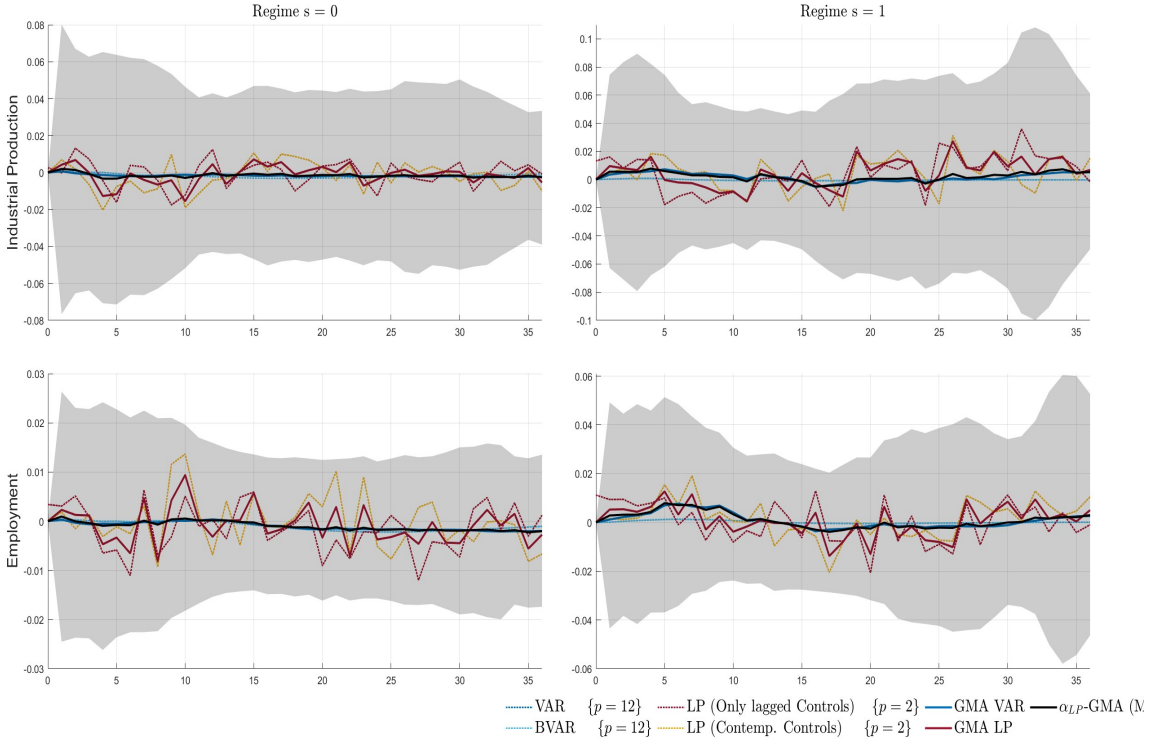
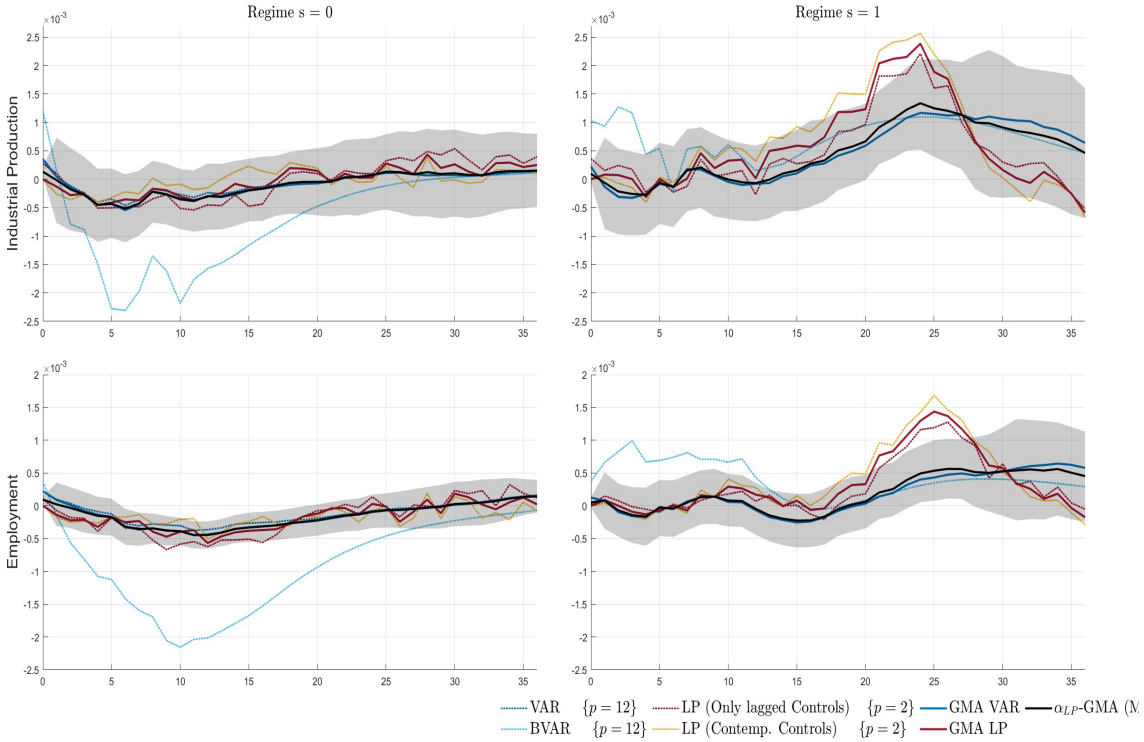


Figure 4

Low vs. High inflation volatility regimes:

The impact of the uncertainty shock



The impact of the monetary shock

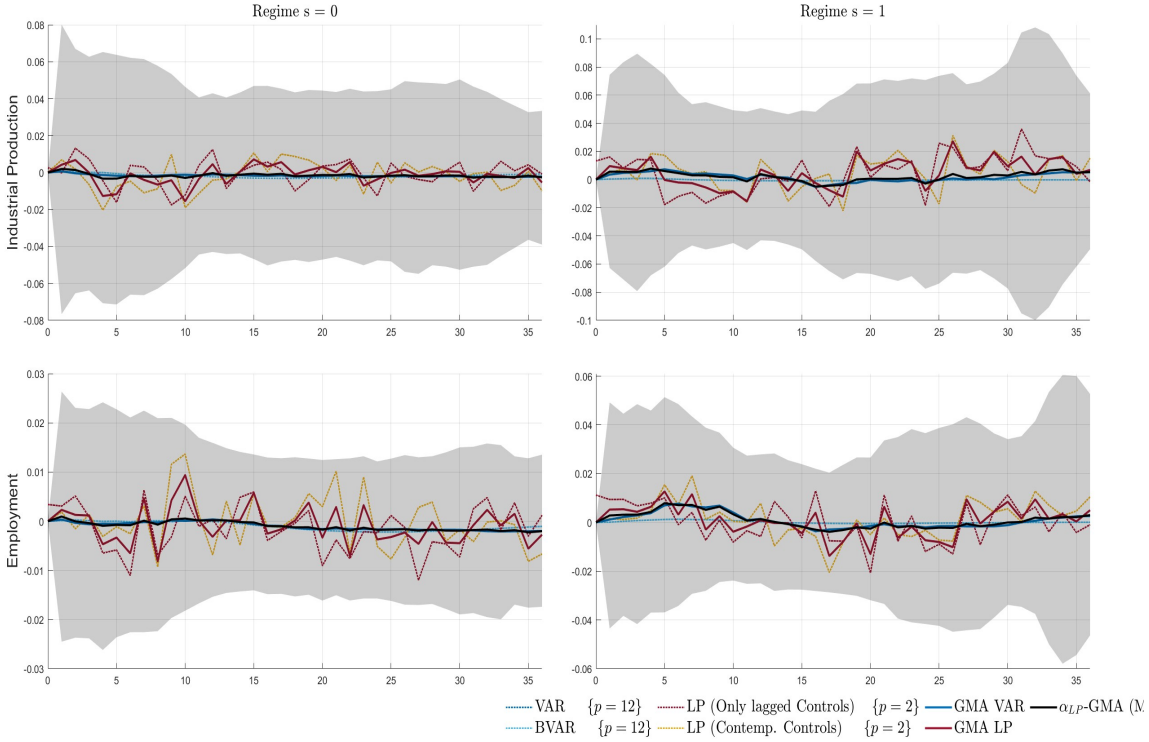


Figure 5