Impulse Response Analysis of Structural Nonlinear Time Series Models

Giovanni Ballarin*
University of Mannheim

March 21, 2024
First version: May 30, 2023

Link to most recent version

Abstract: Linear time series models are the workhorse of structural macroeconometric analysis. Yet, economic theory as well as data suggest that nonlinear and asymmetric effects might be key to understanding the potential effects of sudden economic changes. This paper proposes a new semi-nonparametric sieve approach to estimate impulse response functions of nonlinear time series within a general class of structural models. Using physical dependence conditions, I prove that a two-step procedure can flexibly accommodate nonlinear specifications, avoiding the choice of fixed parametric forms. Sieve impulse responses are proven to be consistent by deriving uniform estimation guarantees, while an iterative algorithm makes it straightforward to compute them in practice. Simulations show that the proposed semi-nonparametric approach provides insurance against misspecification at minor efficiency costs. In a US monetary policy application, I find that the sieve GDP response associated with a rate hike is, at its peak effects, 16% larger than that of a linear model. Finally, when studying interest rate uncertainty shocks, sieve responses imply up to 54% and 71% stronger contractionary effects on production and inflation, respectively.

*E-mail: giovanni.ballarin@gess.uni-mannheim.de. I thank Otilia Boldea, Timo Dimitriadis, Juan Carlos Escanciano, Lyudmila Grigoryeva, Klodiana Istrefi, Marina Khismatullina, So Jin Lee, Yuiching Li, Sarah Mouabbi, Andrey Ramirez, Christoph Rothe, Carsten Trenkler and Mengshan Xu, as well as the participants of the Econonometrics Seminar at the University of Mannheim, the 2023 ENTER Jamboree, the 10th HKMEtrics Workshop, the GSS Weekly Seminar at Tilburg University and the Internal Econometrics Seminar at Vrije Universiteit Amsterdam for their comments, suggestions and feedback.

1 Introduction

This paper presents a semi-nonparametric method to study the structural dynamic effects of unpredictable shocks in a class of nonlinear time series models.

Linear models are the foundation of economic structural time series modeling. The nature of linear models makes them especially tractable and apt at describing fundamental interactions and processes. For example, large classes of macroeconomic models in modern New Keynesian theory can be reduced to linear VARMA form via linearization techniques. This often justifies the application of the linear time series toolbox from a theoretical point of view. Concurrently, the work of Sims (1980) on VARs reinvigorated the strain of macroeconometric literature that seeks to study dynamic economic relationships. Brockwell and Davis (1991), Hamilton (1994b) and Lütkepohl (2005) provide detailed overviews of linear time series modeling and its developments. When the objects of interest are solely dynamic effects, the local projection (LP) approach of Jordà (2005) has also gained popularity as an alternative thanks to its flexibility and ease of implementation. LPs do not directly impose a linear model on the conditional distribution of the time series, but rather consist of linear lag regressions. Throughout this paper, the key dynamic effect under discussion will be the impulse response function (IRF), which is the common inference object of both linear VARMA and LP analyses.

Nonlinear methods seek to flexibly study the dependence structure between variables of interest by accommodating a potentially complex model structure. In recent years, research in nonlinear and asymmetric effects has grown, partly due to the increasing availability of data, making it feasible to estimate more elaborate models (Fuleky, 2020). From a macroeconomic perspective, one can imagine at least three broad categories of nonlinearities that may be important to study. Sign-dependence of impulse responses is a potential key factor in the evaluation of monetary policy, as the specific effects of an interest rate change might be mitigated if the central bank implements a rate drop rather than a rate hike, while some others might be enhanced (Debortoli et al., 2020). If impulse responses are size-sensitive, large shocks and small shocks can have vastly different economic impacts, meaning that the policymaker must account for nonlinear scaling in the intensity of an intervention (Tenreyro and Thwaites, 2016). Finally, if the researcher's objective lies in studying exogenous changes impacting a variable that is nonlinear by definition, such as volatility indexes, any valid structural model should account for this feature.

The main contribution of this work is the development of an approach that allows estimating structural IRFs which can account for general nonlinear effects. This goal entails solving two related issues: first, structural identification of shocks, so that it is possible to give a valid economic interpretation to impulse responses; second, estimation of nonlinear functions in the setting of dependent data. In a linear setup, identification and estimation can be considered as distinct problems, but when working with nonlinear models these questions become intertwined. Without specific assumptions, nonlinear model classes are much too vast in terms of complexity: there are too many channels for any variable to affect any other. Disentangling such channels thus becomes impossible, and one cannot structurally interpret IRFs and dynamic effects such as multipliers. This problem can be solved by being more precise about the classes of models one is willing to entertain. I consider the structural nonlinear framework originally proposed by Gonçalves et al. (2021), which involves selecting one variable to identify the structural shocks of interest, X_t, and treating it separately from all other series, a vector Y_t, included in the model. By imposing a few additional assumptions on the dependence structure of innovations, one is able to include general nonlinear effects of X_t and its lags onto Y_t. By further allowing the lags of Y_t to influence X_t, this setup permits nonlinear dynamics to propagate to all variables over time. The significant upside of this paradigm is that structural identification is built-in, instead of being treated as a separate step. The latter path is most often taken in the literature by implementing the generalized impulse response function (GIRF) proposed by Koop et al. (1996). Kilian and Lütkepohl (2017) have, however, highlighted that common linear identification strategies such as long-term and sign restrictions are generally impossible to impose in general nonlinear models, since closed-form expressions are not available but in a handful of special cases.

A weakness of the framework in Gonçalves et al. (2021) is that it requires choosing a specific functional form for the nonlinear components of the model, such as the negative-censoring map or a cubic map. These are used to tease out the sign and size effects of shocks.[1] Yet, correct prior knowledge of such terms is often unreasonable, especially in multivariate, multi-lag models. The natural way to avoid selecting a parametric nonlinear specification is to resort to semi-nonparametric techniques. Nonparametric time series methods have a long history in econometrics (Härdle et al., 1997), but until recently not much progress has been made in applying them to studying dynamic effects. Impulse response functions are objects that depend on the global properties of the model and, to be more precise, defining an IRF requires iterating shock perturbations over time. In a nonlinear model, the perturbation depends on the variables' state, so that one must consider the shock's effects across possible states. That is, different features of the nonlinear model such as level, slope, curvature must be evaluated over a range of values. Therefore, in this setting, an econometrician must provide error guarantees that are uniform over the variables' domain. In this work, I combine the uniform inference framework of Chen and Christensen (2015) with the structural nonlinear time series scheme discussed above. The general idea is to resort to semi-nonparametric series estimation and work in a physical dependence setup (Wu, 2005). On the one hand, I argue that physical dependence is a natural way of imposing assumptions that lead to estimable models, being more transparent than standard mixing conditions. On the other hand, the series approach makes it easy to estimate models with linear and nonlinear components of the type considered in this paper. It also provides well-developed theoretical results to study uncertainty. Under appropriate regularity assumptions, I show that a two-step semi-nonparametric series estimation procedure is able to consistently recover the structural model in a uniform sense. This result encompasses the generated regressors' problem, which arises in the second step due to the structural identification strategy. Lastly, I prove that the nonlinear impulse response function estimates obtained are themselves asymptotically consistent and, thanks to an iterative algorithm, straightforward to

[1] The negative-censoring map applied to variable a is $a \mapsto \max(a, 0)$.

compute in practice.

To validate the proposed methodology, I provide simulation evidence. The first set of results shows that, with realistic sample sizes, the efficiency costs of the semi-nonparametric procedure are small compared to correctly-specified parametric estimates. A second set of simulations demonstrates that whenever the nonlinear parametric model is mildly misspecified the large-sample bias is large, while for semi-nonparametric estimates it is negligible. Finally, I study how the IRFs computed with the new method compare with the ones from two previous empirical exercises. In a small, quarterly model of the US macroeconomy, I find that the parametric nonlinear and nonlinear appear to underestimate by intensity the GDP responses by 13% and 16%, respectively, after a large exogenous monetary policy shock. Moreover, sieve responses achieve maximum impact a year before their linear counterparts. Then, I evaluate the effects of interest rate uncertainty on US output, prices, and unemployment following Istrefi and Mouabbi (2018). In this exercise, the impact on industrial production of a one-deviation increase in uncertainty is approximately 54% stronger according to semi-nonparametric IRFs than the comparable linear specification. These findings suggest that structural impulse responses predicated on linear specifications might be appreciably underestimating shock effects.

Related Literature. Nonlinear models for dependent data have been extensively developed with the aim of analyzing diverse types of series, see e.g. the monographs of Tong (1990), Fan and Yao (2003), Gao (2007), Tsay and Chen (2018). Teräsvirta et al. (2010) provide a thorough discussion of nonlinear economic time series modeling, but, by only presenting the generalized IRF (GIRF) approach proposed by Koop et al. (1996), Potter (2000) and Gourieroux and Jasiak (2005), they do not explicitly address *structural* analysis.

Parametric nonlinear specifications are common prescriptions, for example, in time-varying models (Auerbach and Gorodnichenko, 2012, Caggiano et al., 2015) and state-depend models (Ramey and Zubairy, 2018). They have been and are commonly used in time-homogeneous models. Kilian and Vega (2011) provide a structural analysis of the effects of GDP on oil price shocks and, in contrast to previous literature, find that asymmetries play a negligible role: they do this by including a negative-censoring transformation of the structural variable and testing for significance. Caggiano et al. (2017), Pellegrino (2021) and Caggiano et al. (2021) use interacted VAR models to estimate effects of uncertainty and monetary policy shocks. From a finance perspective, Forni et al. (2023a,b) study the economic effects of financial shocks. Their generalized VMA specification, which is based on that of Debortoli et al. (2020), sets that innovations be transformed with the quadratic map.[2] Gambetti et al. (2022) study news shocks asymmetries by imposing that news changes enter their autoregressive model with a pre-specified threshold function.

Extension of nonparametric methods to nonlinear time series have already been discussed in the recent literature. For example, Kanazawa (2020) proposed to use radial basis function neural networks to estimate a nonlinear time series model of the US macroeconomy. This work focuses

[2] I will discuss how their nonlinear model setup compares to the one I consider below.

on estimating the GIRF of Koop et al. (1996), with its structural limitations: productivity is assumed to be a fully exogenous variable. Gourieroux and Lee (2023) provide a framework for nonparametric kernel estimation and inference of IRFs via local projections. Yet, they primarily work in the one-dimensional case and only mention economic identification in multivariate setups from the perspective of linear VARs. The work possibly closest to the present paper seems to be that of Lanne and Nyberg (2023), who develop a nearest-neighbor approach to impulse responses estimation that builds on the local projection idea and the GIRF concept. These papers, save for Gourieroux and Lee (2023), do not fully develop an asymptotic theory for their estimators, which makes it hard to judge the econometric assumptions under which they are applicable.

Outline. The remainder of this paper is organized as follows. Section 2 provides the general framework for the structural model. Section 3 describes the two-step semi-nonparametric estimation strategy, provides a thorough treatment of physical dependence assumptions and reports the key uniform consistency guarantees. Section 4 is devoted to the discussion of nonlinear impulse response function computation, validity and consistency. In Section 5, I report simulation results that show the performance of the proposed method, while in Section 6 I discuss empirical applications. Finally, Section 7 concludes. All proofs and additional technical results, as well as secondary plots, can be found in Appendices B and C, respectively.

Notation. A (vector) random variable will be denoted in capital or Greek letters, e.g. Y_t or ϵ_t, while its realization will be in lowercase Latin letters, that is y_t. For a process $\{Y_t\}_{t\in\mathbb{Z}}$, we write $Y_{t:s} = (Y_t, Y_{t+1}, \ldots, Y_{s-1}, Y_s)$, as well as $Y_{*:t} = (\ldots, Y_{t-2}, Y_{t-1}, Y_t)$ for the left-infinite history and $Y_{t:*} = (Y_t, Y_{t+1}, Y_{t+2}, \ldots)$ for its right-infinite history. The same notation is also used for random variable realizations. For a matrix $A \in \mathbb{R}^{d\times d}$ where $d \geqslant 1$, $\|A\|$ is the spectral norm, $\|A\|_\infty$ is the supremum norm and $\|A\|_r$ for $0 < r < \infty$ is the r-operator norm. For a random vector or matrix, I will use $\|\cdot\|_{L^r}$ to denote the associated L^r norm.

2 Model Framework

In this section, I introduce the nonlinear time series model that will be considered throughout the paper. This model setup will be a generalization of the one developed in Gonçalves et al. (2021) by letting the form of nonlinear components to remain unspecified until estimation. The idea behind the partial structural identification scheme is simple: if Z_t is the full vector of time series of interest, one must choose one series, call it X_t, as the structural variable, and add specific assumption on its dynamic effects on the remaining series, vector Y_t. The central goal will be the estimation of the impulse responses of Y_t due to a shock in X_t.

2.1 A Simple Nonlinear Monetary Policy Model

To begin with, it is useful to present a basic modeling setup with a straightforward economic interpretation, which may also serve as a concrete empirical example for the setting I will develop. To this end, I consider first a simple nonlinear monetary policy (MP) model which, however, captures all of the key ingredients of the general framework discussed in the next subsection. Consider the following hypothetical model of US macroeconomic time series:

$$\begin{aligned} X_t &= \rho X_{t-1} + \epsilon_{1t}, \\ \mathrm{FFR}_t &= \alpha_{11}\mathrm{FFR}_{t-1} + \alpha_{12}X_{t-1} + \beta_0^1\epsilon_{1t} + \epsilon_{21t}, \\ \mathrm{GDP}_t &= \alpha_{21}\mathrm{GDP}_{t-1} + \alpha_{22}\mathrm{FFR}_{t-1} + G(X_t) + \beta_0^2\epsilon_{1t} + \epsilon_{22t}, \end{aligned}$$

where X_t is a structural monetary policy variable (for example, a credibility exogenous sequence of autocorrelated MP shocks), FFR_t is the Federal Funds Rate and GDP_t is the US Gross Domestic Product. Moreover, ϵ_{1t} and $\epsilon_{2t} := (\epsilon_{21t}, \epsilon_{22t})'$ are reciprocally independent sequences of shocks. Coefficients α_{11}, α_{12}, α_{21} and α_{22} induce a linear autoregressive structure for the endogenous variables FFR_t and GDP_t, while β_0^1 and β_0^2 determine the structural effects of ϵ_{1t} on FFR and GDP. Moreover, notice that the (sufficiently smooth) nonlinear function $G : \mathbb{R} \to \mathbb{R}$ implies that shocks ϵ_{1t} not only effect GDP contemporaneously in a linear fashion, but also nonlinearly through the level of X_t. To first aid conceptualization, one could think of setting G to be a known transformation which can tease out a specific effect of interest. For example, $G(X_t) = \max(0, X_t)$ to incorporate an asymmetry which depends on the sign of X_t.[3] Yet, a choice of G that is made a priori is hard to justify in general, and so the objective is rather to estimate G jointly with all other parameters. This will be the core issue at hand in the remainder of this paper.

To formally and effectively analyze this simple MP model and discuss its estimation, I separate the linear, nonlinear and structural parts. In vector form,

$$\begin{bmatrix} X_t \\ \mathrm{FFR}_t \\ \mathrm{GDP}_t \end{bmatrix} = \begin{bmatrix} \rho & 0 & 0 \\ \alpha_{12} & \alpha_{11} & 0 \\ 0 & \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ \mathrm{FFR}_{t-1} \\ \mathrm{GDP}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ G(X_t) \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ \beta_0^1 & 1 & 0 \\ \beta_0^2 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{21t} \\ \epsilon_{22t} \end{bmatrix}$$

Now, by setting $Y_t := (\mathrm{FFR}_t, \mathrm{GDP}_t)'$ and $Z_t := (X_t, \mathrm{FFR}_t, \mathrm{GDP}_t)' \equiv (X_t, Y_t')'$, we obtain the equation

$$Z_t = A_1 Z_{t-1} + G_1(X_t) + B_0^{-1}\epsilon_t,$$

where A_1 is a matix function of $(\rho, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$, B_0^{-1} is a matrix function of (β_0^1, β_0^2) and $G_1(X_t) = (0, 0, G(X_t))'$. Throughout this paper, I will call the form in the above display the semi-reduced form, for reasons that will be made clear when presenting the general model. Finally, a key insight is that one can, with a mild abuse of notation, write G_1 too in a "functional matrix" form, that is

$$G_1(X_t) = \begin{bmatrix} 0 \\ 0 \\ G(X_t) \end{bmatrix} X_t \equiv \begin{bmatrix} 0 \\ 0 \\ G \end{bmatrix} X_t = G_1 X_t.$$

Below, this will formalism will prove very useful in terms of streamlining notation.

[3] See also the wider class of threshold autoregressive models (TAR) discussed by Fan and Yao (2003) and Teräsvirta et al. (2010).

2.2 General Model

Let $Z_t := (X_t, Y_t')'$ where $X_t \in \mathcal{X} \subseteq \mathbb{R}$ and $Y_t \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, and let $d = 1 + d_Y$ be the dimension of Z_t. I assume that the structural nonlinear data generating process has the form

$$B_0 Z_t = b + B(L) Z_{t-1} + F(L) X_t + \epsilon_t, \tag{1}$$

where $b = (b_1, b_2')' \in \mathbb{R}^d$ and $\epsilon_t = (\epsilon_1, \epsilon_2')' \in \mathcal{E} \subseteq \mathbb{R}^d$ are partitioned accordingly. Moreover, I assume that model (1) imposes a *linear* dependence of observables on Y_t and its lags, while series X_t can enter *nonlinearly*. That is, $B(L) = B_1 + B_2 L + \ldots + B_p L^{p-1}$ and $F(L) = F_0 + F_1 L + \ldots + F_p L^p$ are linear and functional lag polynomials, respectively.[4]

Matrices $(F_0, \ldots, F_p)$ are functional in the sense that their entries consist of real univariate functions, and the product between $F(L)$ and X_t is to be interpreted as functional evaluation, c.f. the example discussed above. That is,

$$F(L)X_t = \begin{bmatrix} f_{0,1}(X_t) \\ \vdots \\ f_{0,d}(X_t) \end{bmatrix} + \begin{bmatrix} f_{1,1}(X_{t-1}) \\ \vdots \\ f_{1,d}(X_{t-1}) \end{bmatrix} + \ldots + \begin{bmatrix} f_{p,1}(X_{t-p}) \\ \vdots \\ f_{p,d}(X_{t-p}) \end{bmatrix},$$

where $\{f_{j,l}\} \in \Lambda$ for $j = 0, \ldots, p$, $l = 1, \ldots, d$, and Λ is a sufficiently regular function class.[5] The modeling choice to remain within the autoregressive time series class with additive lag structure has two core advantages. First, it yields a straightforward generalization to classical linear models (Lütkepohl, 2005, Kilian and Lütkepohl, 2017). Second, it keeps semi-nonparametric estimation of nonlinear components feasible. Additivity in variables and lags means that the curse of dimensionality involved with multivariate nonparametric estimation is effectively mitigated (Fan and Yao, 2003).

Let the lag polynomials be given by

$$B(L) = \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix}, \qquad F(L) = \begin{bmatrix} 0 \\ F_{21}(L) \end{bmatrix}.$$

This structural formulation means that the model equation for X_t is restricted to be linear in all regressors. It also implies that X_t does not depend contemporaneously on itself. Note that as long as $B_{12}(L) \neq 0$, X_t still depends upon nonlinear functions of its own lags, which enter via lags of Y_t. Next, I impose that $B_0 \in \mathbb{R}^{d_Y \times d_Y}$ has the form

$$B_0 = \begin{bmatrix} 1 & 0 \\ -B_{0,12} & B_{0,22} \end{bmatrix},$$

[4]This is a minor abuse of notation compared to e.g. Lütkepohl (2005). The choice to use a matrix notation is due to the ease and clarity of writing a (multivariate) additive nonlinear model such as (1) in a manner consistent with standard linear VAR models. In cases where a real matrix $A \in \mathbb{R}^{d \times d}$ is multiplied with a conformable functional matrix F, I simply assume the natural product of a scalar times a function, e.g. $A_{ij} F_{k\ell}$, where $F_{k\ell}$ is a function, returning a new real function.

[5]To fix ideas, one may think of $\Lambda^q(M)$, the Hölder function class of smoothness $q > 0$ and domain $M \subseteq \mathbb{R}$. We shall make more precise assumptions regarding Λ in Section 3 when discussing model estimation.

where $B_{0,22}$ is non-singular and normalized to have unit diagonal. The structural model is thus given by

$$
\begin{aligned}
X_t &= b_1 + B_{12}(L)Y_{t-1} + B_{11}(L)X_{t-1} + \epsilon_{1t}, \\
B_{0,22}\, Y_t &= b_2 + B_{22}(L)Y_{t-1} + B_{21}(L)X_{t-1} + B_{0,12}X_t + F_{21}(L)X_t + \epsilon_{2t}.
\end{aligned}
$$

Moreover, it follows that B_0^{-1} exists and has form

$$
B_0^{-1} = \begin{bmatrix} 1 & 0 \\ B_0^{21} & B_0^{22} \end{bmatrix}.
$$

The constraints on B_0 yield a structural identification assumption and require that X_t be pre-determined with respect to Y_t (Gonçalves et al., 2021). By introducing

$$
\mu := B_0^{-1}b, \quad A(L) := B_0^{-1}B(L) \quad \text{and} \quad G(L) := B_0^{-1}F(L),
$$

one thus obtains

$$
\begin{aligned}
X_t &= \mu_1 + A_{12}(L)Y_{t-1} + A_{11}(L)X_{t-1} + \epsilon_{1t}, \\
Y_t &= \mu_2 + A_{22}(L)Y_{t-1} + A_{21}(L)X_{t-1} + G_{21}(L)X_t + B_0^{21}\epsilon_{1t} + B_0^{22}\epsilon_{2t},
\end{aligned}
\tag{2}
$$

or, equivalently,

$$
Z_t = \mu + A(L)Y_{t-1} + G(L)X_t + u_t, \tag{3}
$$

where $u_t = [u_{1t}, u_{2t}]'$, $u_{1t} \equiv \epsilon_{1t}$ and $u_{2t} := B_0^{21}\epsilon_{1t} + B_0^{22}\epsilon_{2t}$. Given the structure of B_0^{-1}, one can see that $A_{12}(L) \equiv B_{12}(L)$, $A_{11}(L) \equiv B_{11}(L)$ and $G_{11}(L) = 0$. Importantly, one must also notice that $A_{12}(L)$ and $G_{21}(L) = B_0^{22}F_{21}(L)$ might now be not properly identified without further assumptions. Since $A_{21}(L)$ is not necessarily zero, linear effects of lags of X_t on Y_t can enter by means of both lag polynomials. To resolve this issue, I therefore assume that the functional polynomial $G_{21}(L)$ contains, at lags greater than zero, only *nonlinear* components.[6]

Example 2.1. (Bivariate Model with Exogenous Shocks). To give a concrete example of (2), assume that one wants to model the effects of monetary policy shocks on U.S. GDP growth following Romer and Romer (2004). Then, let

$$
\begin{aligned}
X_t &= \epsilon_{1t}, \\
Y_t &= \mu_2 + A_2 Y_{t-1} + G(X_t) + B_0^{21}\epsilon_{1t} + \epsilon_{2t},
\end{aligned}
$$

where X_t are the policy shocks, which are assumed to be i.i.d., while Y_t is a macroeconomic variable whose responses the researcher is interested in, e.g. GDP growth or PCE inflation. This setup is very minimal, and I assume here, for the sake of simplicity, that endogeneity of ϵ_{2t} does not pose a problem. Then, the term $G(X_t) + B_0^{21}\epsilon_{1t} \equiv G(\epsilon_{1t}) + B_0^{21}\epsilon_{1t} =: H(\epsilon_{1t})$ fully

[6] When using a semi-nonparametric estimation strategy with B-splines, this will be feasible to implement numerically. When using wavelets, this also is a natural approach. In practice, however, some care must be taken to avoid constructing collinear regression matrices.

captures any contemporaneous effect of monetary policy shocks on Y_t. When $G(\epsilon_{1t}) = 0$, $H(\epsilon_{1t})$ and the model are linear. If $G(\epsilon_{1t}) = \beta_0 \max(0, \epsilon_{1t})$ for some $\beta_0 \neq 0$, function H is piece-wise linear: contractionary and expansionary shocks have, in general, different effects on Y_t, but shocks with the same sign have proportional impact. As a final example, if $G(\epsilon_{1t}) = \beta_0\, \epsilon_{1t}^3$ then $H(\epsilon_{1t})$ is a third-degree polynomial, so that both sign and size of monetary policy shocks are fundamental determinants of Y_t's impulse response function. In principle, to correctly quantify the repercussions of a specific monetary intervention a researcher must model all of these effects, unless they have a strong prior belief that either or both can be safely ignored. More complex nonlinear and asymmetric relations are also possible. A more robust strategy - as proposed in the present work - is to avoid choosing G (or H) as part of the model's specification, but rather to empirically estimate it jointly with all other coefficients.

Remark 2.1. (Constrained Models). The general approach of leaving $F(L)$ unconstrained is appealing when no precise economic intuition or information is available. However, there might be cases where the functional form of the nonlinear component is either partially known, or can be restricted. A simple restriction is that of a uniform functional over lags,

$$F(L) = F + FL + FL^2 + \ldots + FL^p.$$

This is a constraint effectively imposed by e.g. Gonçalves et al. (2021), Kilian and Vega (2011) and other references. They do this by fully specifying F, but nonparametric constraints may be desired, e.g. monotonicity. Constrained estimation of $F(L)$ is addressed in Remark 3.2 below.

The system of equations in (2) provides the so-called *pseudo-reduced form* model. By design, one does not need to identify the model fully, meaning that fewer assumptions on Z_t and ϵ_t are needed to estimate the structural effects of ϵ_{1t} on Y_t. This comes at the cost of not being able to simultaneously study structural effects with respect to ϵ_{2t}. An associated problem is that, in general, $G_{21}(L)X_t$ is correlated with innovation u_{2t} through $B_0^{21}\epsilon_{1t}$. The main challenge to structural shock identification of ϵ_{1t} thus lies in the fact that if $B_0^{21} \neq 0$ and $G_{21}(0) \neq 0$, there is endogeneity in the equations for Y_t since X_t depends linearly on ϵ_{1t}. Gonçalves et al. (2021) address the issue by proposing a two-step estimation procedure wherein one explicitly controls for ϵ_{1t} by using regression residuals $\widehat{\epsilon}_t$. In Section 3 below, I show that this approach also allows for consistent semi-nonparametric estimation of structural impulse responses.

Remark 2.2. (Identification Schemes). Forni et al. (2023a,b) provide an alternative nonlinear structural identification framework to that of Gonçalves et al. (2021). Their approach was originally introduced in Debortoli et al. (2020) and is based on the VMA form of the time series. Using the current notation, suppose that the structural representation of Z_t is given by

$$Z_t = b + Q(L)F(\epsilon_{1t}) + B(L)\epsilon_t$$

where ϵ_t are independent structural shocks with zero mean and identity covariance, while ϵ_{1t} identifies, e.g., financial innovations and shocks. $Q(L)$ and $B(L)$ are both linear lag polynomials and F is a nonlinear function to be specified by the researcher. Imposing some additional assumptions, the reduced-form assumed by Forni et al. (2023a) is

$$Z_t = \mu + A(L)Z_t + Q_0 F(\epsilon_{1t}) + B_0\epsilon_t, \tag{4}$$

where $F(x) = x^2$ in their baseline specification. Forni et al. (2023b) use an analogous model, while Debortoli et al. (2020) also consider more general setups where Q_0 is replaced by a general lag polynomial $D(L)$. These kinds of structural assumptions are similar but not identical to the ones imposed in Gonçalves et al. (2021) and this paper. For (4) to overlap with (2), one must assume that X_t is exogenous and independently distributed, so that its level does not affect the mapping of ϵ_{1t} through F. That is, (4) requires that only the *shocks* have nonlinear effects, not the structural variable itself. The upside of this approach is that one can directly and explicitly model asymmetry in the innovation process. The drawbacks are that, without a clear identification of a structural variable, one must fully identify B_0. Moreover, function F remains to be specified a priori. Note, however, that if innovation sequence ϵ_{1t} is observable, a generalization of the semi-nonparametric estimation results of this paper to the framework of Debortoli et al. (2020) would be straightforward.

I now state some preliminary assumptions for the model.

Assumption 1. $\{\epsilon_{1t}\}_{t\in\mathbb{Z}}$ and $\{\epsilon_{2t}\}_{t\in\mathbb{Z}}$ are mutually independent time series such that

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} \left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \Sigma_2 \end{bmatrix}\right)$$

where Σ_2 is a diagonal positive definite matrix.

Assumption 2. $\{Z_t\}_{t\in\mathbb{Z}}$ is strictly stationary, ergodic and such that $\sup_t \mathbb{E}[|Z_t|] < \infty$.

Assumption 3. The roots of equation $\det(I_d - A(L)L) = 0$ are outside the complex unit circle.

Assumption 1 follows Gonçalves et al. (2021). Assumption 2 is a high-level assumption on the properties of process $\{Z_t\}_{t\in\mathbb{Z}}$ and is common in the analysis of structural time series. Assumption 3 ensures that it is possible to invert lag polynomial $(I - A(L)L)$ in order to define impulse responses, as done below. However, Assumption 2 and 3 will not be sufficient to make sure that (2) is estimable from data, and in Section 3 additional constraints on $A(L)$ and $G(L)$ will be required in order to apply semi-nonparametric estimation. Moreover, Assumption 2 is not easily interpretable: functional lag polynomial $G(L)$ makes it impossible to reduce semi-structural equations (2) to an explicit infinite moving average form.

I will resolve both the former (sufficiency) and latter (interpretability) issue by using the nonlinear dynamic model framework outlined by Pötscher and Prucha (1997). It will allow introducing regularity assumptions on the dependence of Z_t which enable the derivation of consistency of impulse response estimates.

2.3 Structural Nonlinear Impulse Responses

Starting from pseudo-reduced equations (2), by letting $\Psi(L) = (I_d - A(L)L)^{-1}$ one can further derive that

$$Z_t = \eta + \Theta(L)\epsilon_t + \Gamma(L)X_t, \tag{5}$$

where

$$\mu := \Psi(1)\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \quad \Theta(L) := \Psi(L)B_0^{-1}, \quad \text{and} \quad \Gamma(L) := \Psi(L)\begin{bmatrix}0 \\ G_{21}(L)\end{bmatrix}.$$

To formally define impulse responses, it is useful to partition the polynomial $\Theta(L)$ according to

$$\Theta(L) := \begin{bmatrix}\Theta_{\cdot 1}(L) & \Theta_{\cdot 2}(L)\end{bmatrix},$$

where $\Theta_{\cdot 1}(L)$ represents the first column of matrices in $\Theta(L)$, and $\Theta_{\cdot 2}(L)$ the remaining d_Y columns.

Given impulse $\delta \in \mathbb{R}$ at time t, define the shocked innovation process as $\epsilon_{1s}(\delta) = \epsilon_s$ for $s \neq t$ and $\epsilon_{1t}(\delta) = \epsilon_t + \delta$, as well as the shocked structural variable as $X_s(\delta) = X_t$ for $s < t$ and $X_s(\delta) = X_s(Z_{t-1}, \epsilon_t + \delta, \epsilon_{t+1} \ldots, \epsilon_s)$ for $s \geqslant t$. Further, let

$$\begin{aligned} Z_{t+h} &:= \eta + \Theta_{\cdot 1}(L)\epsilon_{1t+h} + \Theta_{\cdot 2}(L)\epsilon_{2t+h} + \Gamma(L)X_t, \\ Z_{t+h}(\delta) &:= \eta + \Theta_{\cdot 1}(L)\epsilon_{1t+h}(\delta) + \Theta_{\cdot 2}(L)\epsilon_{2t+h} + \Gamma(L)X_t(\delta), \end{aligned}$$

be the time-t baseline and shocked series, respectively. The unconditional impulse response is given by

$$\mathrm{IRF}_h(\delta) = \mathbb{E}\left[Z_{t+h}(\delta) - Z_{t+h}\right]. \tag{6}$$

The difference between shock and baseline is clearly

$$\begin{aligned} Z_{t+h}(\delta) - Z_{t+h} &= \Theta_{h,\cdot 1}\delta + \Gamma(L)X_t(\delta) - \Gamma(L)X_t \\ &= \Theta_{h,\cdot 1}\delta + \left(\Gamma_0 X_{t+h}(\delta) - \Gamma_0 X_{t+h}\right) + \ldots + \left(\Gamma_h X_t(\delta) - \Gamma_h X_t\right), \end{aligned}$$

therefore the unconditional IRF reduces to

$$\mathrm{IRF}_h(\delta) = \Theta_{h,\cdot 1}\delta + \mathbb{E}\left[\Gamma_0 X_{t+h}(\delta) - \Gamma_0 X_{t+h}\right] + \ldots + \mathbb{E}\left[\Gamma_h X_t(\delta) - \Gamma_h X_t\right]. \tag{7}$$

Notice that, in (7), while one can linearly separate expectations in the impulse response formula, terms $\mathbb{E}\left[\Gamma_j X_{t+j}(\delta) - \Gamma_j X_{t+j}\right]$ for $0 \leqslant j \leqslant h$ cannot be meaningfully simplified. Coefficients Γ_j are functional, therefore it is not possible to collect them across $X_{t+j}(\delta)$ and X_{t+j}. Moreover, these expectations involve nonlinear functions of lags of X_t and cannot be computed explicitly. To address this issue, Section 4 provides an iterative procedure that makes computation of nonlinear impulse responses in (7) straightforward.

Remark 2.3. (Local Projection Approaches). As mentioned in the introduction, in recent years there has been growing interest in nonlinear IRF estimation procedures, and, accordingly, ways to generalize the LP framework. Jordà (2005) already suggested that nonlinear impulse responses can, in principle, be directly estimated with local projections via the so-called *flexible local projection* approach. The flexible LP method relies on the Volterra expansion of time series to account for nonlinearities. There are multiple issues with this method. First, Jordà (2005) does not directly state how the validity of Volterra series implies the autoregressive form used in the LP regression. Second, the flexible LP proposal is fundamentally equivalent to adding polynomial factors to the linear regression specification. Thus, it is effectively a semi-nonparametric method, yet Jordà (2005) does not provide a theoretical analysis from this viewpoint. Moreover, no criterion or empirical rule-of-thumb for selecting the truncation order of the Volterra expansion are suggested, which becomes a key issue in practice. Due to these concerns, application of flexible LPs seems hard to justify from an econometric perspective.[7] Lanne and Nyberg (2023) propose to nonparametrically recover the conditional mean function with a nearest-neighbor (k-NN) regression estimator. Their method is very flexible, but requires appropriately choosing the neighborhood size k and a distance measure for histories of realizations, and the authors do not theoretically address these issues. Very recently, Gourieroux and Lee (2023) have considered nonlinear IRF estimation with kernel-based methods by means of a novel conditional quantile representation of the process. They prove kernel LP estimators based on such representation are consistent, and that the direct estimator is asymptotically normal. The theory is developed only for the univariate case, with an autoregressive structure of lag order one, limiting the applicability of their procedure.

3 Estimation

Pseudo-reduced form model (2) can be compactly rewritten as

$$\begin{aligned} X_t &= \Pi_1' W_{1t} + \epsilon_{1t}, \\ Y_t &= \Pi_2' W_{2t} + u_{2t}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} \Pi_1 &:= \left(\eta_1,\ A_{1,11},\ \cdots,\ A_{p,11},\ A_{1,12}',\ \cdots,\ A_{p,12}'\right)' \in \mathbb{R}^{1+pd}, \\ \Pi_2 &:= \begin{bmatrix}\eta_2 & G_{1,21} & \cdots & G_{p,21} & A_{1,22} & \cdots & A_{p,22} & B_0^{21}\end{bmatrix}', \\ Z_{t-1:t-p} &:= \left(X_{t-1},\ \ldots,\ X_{t-p}, Y_{t-1}',\ \ldots,\ Y_{t-p}'\right)' \in \mathbb{R}^{pd}, \\ W_{1t} &:= \left(1,\ Z_{t-1:t-p}'\right)' \in \mathbb{R}^{1+pd}, \\ W_{2t} &:= \left(1,\ X_t, Z_{t-1:t-p}',\ \epsilon_{1t}\right)' \in \mathbb{R}^{3+pd}. \end{aligned}$$

Additionally, let $W_1 = (W_{11}, \ldots, W_{1n})'$ and $W_2 = (W_{21}, \ldots, W_{2n})'$ be the design matrices for X_t and Y_t, respectively.

[7] Moreover, the complexity of estimating Volterra kernels grows exponentially with the kernel order, and thus more sophisticated approaches have been proposed to make estimation feasible, see e.g. Sirotko-Sibirskaya et al. (2020) and Movahedifar and Dickhaus (2023).

Two-step Estimation Procedure. Since W_{2t} is an infeasible vector of regressors, to estimate Π_2 one can use $\widehat{W}_{2t} = (1, X_t, Z'_{t-1:t-p}, \widehat{\epsilon}_{1t})'$, which now contains generated regressors in the form of residual $\widehat{\epsilon}_{1t}$. This approach is an adaptation of the two-step procedure put forth by Gonçalves et al. (2021), where I allow for semi-nonparametric estimation:

1. Regress X_t onto W_{1t} to get estimate $\widehat{\Pi}_1$ and compute residuals $\widehat{\epsilon}_{1t} = X_t - \widehat{\Pi}_1' W_{1t}$.

2. Fit Y_t using $\widehat{W}_{2t}$ to get estimate $\widehat{\Pi}_2$. Since $G_{1,21}, \ldots, G_{p,21}$ contain functional parameters, a semi-nonparametric estimation method is required.

3. Compute coefficients in $\widehat{\Theta}(L)$ and $\widehat{\Gamma}(L)$ from $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$.

4. Consider the two paths with time t shocks $\epsilon_t + \delta$ versus ϵ_t: to construct the unconditional IRF, average over histories as well as future shocks by using the algorithm detailed in Proposition 4.1 or Proposition 4.2.

Gonçalves et al. (2021) only allow for pre-determined nonlinear transforms of X_t. The core contribution of this paper is allowing $G_{1,21}, \ldots, G_{p,21}$ to be estimated in a nonparametric way. I focus on series estimation in order to build on the extensive theory available in the setting of dependent data (Chen, 2013, Chen and Christensen, 2015). This further adds to the framework of Gonçalves et al. (2021), as their regularity assumptions are stated only as preconditions for a uniform LLN to hold and are not easy to interpret.

Remark 3.1. (Alternative Estimation Approaches). One does not need to limit estimation of the nonlinear functional parameters $G_{1,21}, \ldots, G_{p,21}$ to series-type estimators. The literature on nonparametric regression is mature, and thus kernel (Tsybakov, 2009), nearest-neighbor (Li and Racine, 2009), partitioning (Cattaneo et al., 2020) and deep neural network (Farrell et al., 2021) estimators are all potentially valid alternatives. For example, Huang et al. (2014) use kernel regression to perform density estimation and regression under physical dependence. However, thanks to both availability of uniform inference results (see also Belloni et al. 2015) and ease of implementation, series methods stand out as a choice for semi-nonparametric time series estimation and nonlinear impulse response computation.

In the reminder of this section, I first introduce the semi-nonparametric series estimation strategy in detail. Then, I outline the core assumptions of the sieve setup. Special focus is put on the dependence structure of the data: rather than directly assuming β-mixing as in Chen and Christensen (2015), I shall consider physical dependence assumptions (Wu, 2005). to provide transparent conditions on the model itself that, if satisfied, ensure consistency. I prove that the proposed two-step semi-nonparametric procedure is uniformly consistent under physical dependence assumptions. These assumptions can be imposed directly on the model, and, as such, may be empirically checked, if necessary. The uniform asymptotic guarantees are first stated for the infeasible estimator involving true innovations ϵ_{1t} and later extended to encompass feasible estimator $\widehat{\Pi}_2$.

3.1 Semi-nonparametric Series Estimation

Starting from (8), one can introduce the ith-row coefficient matrices

$$\begin{aligned} G_i^{21} &= \begin{bmatrix} G_{1,21} & \cdots & G_{p,21} \end{bmatrix}_i , \\ A_i^{22} &= \begin{bmatrix} A_{1,22} & \cdots & A_{p,22} \end{bmatrix}_i , \end{aligned}$$

and B_{0i}^{21} accordingly. Consider now the regression problem for each individual component of Y_t,

$$Y_{it} = G_i^{21} X_{t:t-p} + A_i^{22} Y_{t-1:t-p} + B_{0i}^{21} \epsilon_{1t} + u_{2it},$$

where $X_{t:t-p} := (X_t, \ldots, X_{t-p})'$ and $i = 1, \ldots, d_Y$. For simplicity of notation, I suppress intercept η_{2i}, but this is without loss of generality. Since G_i^{21} consists of $1+p$ functional coefficients and A_i^{22} can be segmented into p row vectors of length d_Y, it is possible to rewrite the above as

$$Y_{it} = \sum_{j=0}^{p} g_{ij}^{21}(X_{t-j}) + \sum_{j=1}^{p} A_{ij}^{22}\, Y_{t-j} + B_{0i}^{21} \epsilon_{1t} + u_{2it}. \tag{9}$$

I will use $\pi_{2,i} := [G_i^{21}\ A_i^{22}\ B_{0i}^{21}]'$ to identify the vector of coefficients in the equation for the ith component of Y_t. From (9), Π_2' can be decomposed in d_Y rows of coefficients, i.e.

$$\begin{bmatrix} Y_{1t} \\ \vdots \\ Y_{d_Y t} \end{bmatrix} = \begin{bmatrix} \pi_{2,1} \\ \vdots \\ \pi_{2,d_Y} \end{bmatrix} W_{2t} + u_{2t}$$

and one can treat each equation separately.

A semi-nonparametric series estimator for (9) is built on the idea that, if functions g_{ij}^{21} belong to an appropriate functional space, one can construct a growing collection of sets of basis functions – called a *sieve* – which, linearly combined, progressively approximate g_{ij}^{21}. That is, one can reduce the infinite dimensional problem of estimating the functional coefficients in $\pi_{2,i}$ to a linear regression problem. Although (9) features a sum of possibly nonlinear functions in $\{X_{t-j}\}_{j=0}^{p}$, as well as linear functions of $\{Y_{t-j}\}_{j=1}^{p}$ and ϵ_{1t}, constructing a sieve is straightforward.[8]

Assume that $g_{ij}^{21} \in \Lambda$, where Λ is a sufficiently regular function class to be specified in the following, and let $\mathbf{B}_\Lambda$ be a sieve for Λ. Let $b_{1\kappa}, \ldots, b_{\kappa\kappa}$ be the collection of $\kappa \geqslant 1$ sieve basis functions from $\mathbf{B}_\Lambda$ and define

$$\begin{aligned} b^\kappa(x) &:= (b_{1\kappa}(x), \ldots, b_{\kappa\kappa}(x))' , \\ B_\kappa &:= (b^\kappa(X_{1:1-p}), \ldots, b^\kappa(X_{n:n-p}))' . \end{aligned}$$

The sieve space for $\pi_{2,i}$ is $\mathbf{B}_\Lambda^{1+p} \times \mathbb{R}^{1+pd_Y}$, where here $\mathbb{R}$ identifies the space of linear functions. Since the nonparametric components of Π_2 are linearly separable in the lag dimension, I take

[8] See Chen (2007) for a comprehensive exposition of sieve estimation. Chen and Shen (1998) and Chen (2013) also provide additional examples of partially linear semi-nonparametric models under dependence.

$\mathbf{B}_{\Lambda}^{1+p}$ to be a direct product of sieve spaces.[9] Importantly, the same sieve can be used for all components of Y_t, as I assume the specification of the model does not change across i.

Let $b_{\pi,1K}, \dots, b_{\pi,KK}$ be the sieve basis in $\mathbf{B}_{\Lambda}^{1+p} \times \mathbb{R}^{1+pd_Y}$ which, for $\kappa \geqslant 1$ and $K = p\kappa + (1 + pd_Y)$, is given by

$$\begin{aligned} b_{\pi,1K}(W_{2t}) &= b_{1\kappa}(X_t), \\ &\vdots \\ b_{\pi,(p\kappa)K}(W_{2t}) &= b_{\kappa\kappa}(X_{t-p}), \\ b_{\pi,(p\kappa+1)K}(W_{2t}) &= Y_{t-1,1}, \\ &\vdots \\ b_{\pi,(K-1)K}(W_{2t}) &= Y_{t-p,d_Y}, \\ b_{\pi,KK}(W_{2t}) &= \epsilon_{1t}, \end{aligned}$$

where κ fixes the size of the nonparametric component of the sieve. Note that K, the overall size of the sieve, grows linearly in κ, which itself controls the effective dimension of the nonparametric component of the sieve, $b_{\pi,1\kappa}, \dots, b_{\pi,\kappa\kappa}$. In all theoretical results, I will focus on the growth rate of K rather than κ, as asymptotically they differ at most by a constant multiplicative factor.

The regression equation for $\pi_{2,i}$ is

$$Y_i = \pi_{2,i}' W_2 + u_{2i},$$

where $Y_i = (Y_{i1}, \dots, Y_{in})'$ and $u_{2i} = (u_{2i1}, \dots, u_{2in})'$. The estimation target is the conditional expectation $\pi_{2,i}(w) = \mathbb{E}[Y_{it} \mid W_{2t} = w]$ under the assumption $\mathbb{E}[u_{2it} \mid W_{2t}] = 0$. By introducing

$$\begin{aligned} b_\pi^K(w) &:= (b_{\pi,1K}(w), \dots, b_{\pi,KK}(w))', \\ B_\pi &:= \left(b_\pi^K(W_{21}), \dots, b_\pi^K(W_{2n})\right)', \end{aligned}$$

the *infeasible least squares series estimator* $\widehat{\pi}_{2,i}^*(w)$ is given by

$$\widehat{\pi}_{2,i}^*(w) = b_\pi^K(w)'(B_\pi' B_\pi)^{-1} B_K' Y_i.$$

Similarly, consider the feasible series regression matrices

$$\begin{aligned} b_\pi^K(w) &:= (b_{\pi,1K}(w), \dots, b_{\pi,KK}(w))', \\ \widehat{B}_\pi &:= \left(b_\pi^K(\widehat{W}_{21}), \dots, b_\pi^K(\widehat{W}_{2n})\right)'. \end{aligned}$$

[9] It is not necessary to consider the more general case of tensor products of 1D sieve functions, as it would be the case for a general $(1 + d_Y)$-dimensional function $G_i^{21}(X_t, X_{t-1}, \dots, X_{t-p})$. As previously discussed, the additive structure avoids the curse of dimensionality which in nonlinear time series modeling if often a primary concern when working with moderate sample sizes (Fan and Yao, 2003).

Thus, the *feasible least squares series estimator* is

$$\widehat{\pi}_{2,i}(w) = b^K_\pi(w)'(\widehat{B}'_\pi \widehat{B}_\pi)^{-1}\widehat{B}'_K Y_i.$$

Given that the semi-nonparametric estimation problem is the same across i, to further streamline notation, where it does not lead to confusion I will let π_2 be a generic coefficient vector belonging to $\{\pi_{2,i}\}_{i=1}^p$, as well as define $\widehat{\pi}_2$, Y and u_2 accordingly.

Remark 3.2. (Constrained Sieve Estimation). The idea of constrained estimation was only briefly touched upon in Remark 2.1. In fully parametric nonlinear models, constraints are often imposed out of necessity or simplicity. If, say, $G_{1,21}$ is constituted only of the negative-censoring map, it is unclear why $G_{2,21}$ would be constituted instead of quadratic or cubic functions, for example. That is, *specific* parametric assumptions can be either unreasonable or hard to justify in practice.[10] Yet, constrained semi-nonparametric estimation might be desirable at times.

If the shape of the regression function is to be constrained to ensure e.g. non-negativity, monotonicity or convexity, Chen (2007) gives examples of shape-preserving sieves, like cardinal B-spline wavelets. Constraints on a generic sieve can also be imposed at estimation time. For example, for simplicity suppose $d_Y = 1$ and $p = 2$, and that one wants to impose $G_{1,21} = G_{2,21}$. The constrained sieve estimator then solves

$$\min_{\beta} \sum_{t=p+1}^{n} \big(Y_t - \beta' b^K_\pi(W_{2t})\big)^2 \quad \text{subject to} \quad \big[I_\kappa, -I_\kappa, 0_{\kappa\times(1+pd_Y)}\big]\beta = 0.$$

Analysis of restricted or constrained estimators, however, is still a challenging problem in nonparametric theory, c.f. Horowitz and Lee (2017), Freyberger and Reeves (2018), Chetverikov et al. (2018). Misspecification in particular is complex to address. Accordingly, I will not be imposing any specific restrictions on the nonlinear functions in Π_2 outside the ones necessary to derive uniform asymptotic theory.

Spline Sieve. The B-spline sieve $\mathrm{BSpl}(\kappa, [0,1]^{d_Y}, r)$ of degree $r \geqslant 1$ over $[0,1]^{d_Y}$ can be constructed using the Cox-de Boor recursion formula. Alternatively, an equivalent way of constructing the spline sieve is as follows. For simplicity, let $d_Y = 1$ and let $0 < m_1 < \ldots < m_{\kappa-r-1} < 1$ be a set of knots. Then

$$b^\kappa_{\text{spline}}(x) := \big(1, x, x^2, \ldots, x^r, \max(x - m_1, 0)^r, \ldots, \max(x - m_{\kappa-r-1}, 0)^r\big)'.$$

The resulting spline sieve is piece-wise polynomial of degree r. Moreover, notice that in practice the spline sieve already contains a linear and constant term, so care must be taken to avoid collinearity (for example, by not including an additional intercept and linear term in X_t in the series regression).

[10]For more precise examples and a more in-depth discussion, see Section 2.1 of Chen (2013).

3.2 Distributional and Sieve Assumptions

To develop the asymptotic uniform consistency theory, I rely on the general theoretical framework established by Chen and Christensen (2015). Basic distributional and sieve assumptions can be carried over from their setup mostly unchanged.

Assumption 4. (i) $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are such that $\epsilon_t \overset{\text{i.i.d.}}{\sim} (0, \Sigma)$, (ii) $\{\epsilon_{1t}\}_{t\in\mathbb{Z}}$ and $\{\epsilon_{2t}\}_{t\in\mathbb{Z}}$ are mutually independent, (iii) $\epsilon_t \in \mathcal{E}$ for all $t \in \mathbb{Z}$ where $\mathcal{E} \subset \mathbb{R}^{d_Y}$ is compact, convex and has nonempty interior.

Assumption 5. (i) $\{Z_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic time series, (ii) $X_t \in \mathcal{X}$ for all $t \in \mathbb{Z}$ where $\mathcal{X} \subset \mathbb{R}$ is compact, convex and has nonempty interior, (iii) $Y_t \in \mathcal{Y}$ for all $t \in \mathbb{Z}$ where $\mathcal{Y} \subset \mathbb{R}^{d_Y}$ is compact, convex and has nonempty interior.

Assumptions 4(i)-(ii) are a repetition of Assumption 1. As W_{2t} depends only on $X_{t:t-p}$, $Y_{t-1:t-p}$, and ϵ_{1t}, Assumption 1 also implies that entries of u_{2t} are independent of W_{2t}, so that $\mathbb{E}[u_{2it} \mid W_{2t}] = 0$.[11] Assumption 5(i) also follows from Assumption 2. However, thanks to the results derived in Section 3.3, below I will impose more primitive conditions on the model for Z_t that allow to recover 5(i). Assumption 4(iii) and Assumptions 5(ii)-(iii) imply that X_t, Y_t, as well as ϵ_t are bounded random variables. In (semi-)nonparametric estimation, imposing that X_t be bounded almost surely is a standard assumption. Since lags of Y_t and innovations ϵ_t contribute linearly to all components of Z_t, it follows that they too must be bounded. Unbounded regressors are more complex to handle when working in the nonparametric setting. Generalization from bounded to unbounded domains under dependence has already been discussed by e.g. Fan and Yao (2003). Chen and Christensen (2015) also allow for an expanding support by using weighted sieves. I leave this extension for future work.

It is, however, important to highlight that bounded support assumptions are relatively uncommon in time series econometrics. This is clear when considering the extensive literature available on linear models such as, e.g., state-space, VARIMA and dynamic factor models (Hamilton, 1994a, Lütkepohl, 2005, Kilian and Lütkepohl, 2017, Stock and Watson, 2016). Avoiding Assumptions 4(iii) and 5(iii) can possibly be achieved with a change in the model's equations – so that, for example, lags of Y_t only effect X_t either via bounded functions or not at all – so I do not discuss this approach here. In practice, Assumptions 4(ii) and 5(ii)-(iii) are not excessively restrictive, as most credibly stationary economic series often have reasonable implicit (e.g. inflation) or explicit bounds (e.g. employment rate).[12]

Let $\mathcal{F}_t = \sigma(\ldots, \epsilon_{1t-1}, u_{2t-1}, Y_{t-1}, \epsilon_{1t}, u_{2t}, Y_t)$ be the natural filtration defined up to time t. Thanks to Assumptions 4 and 5 the following moment requirements hold trivially.

[11] Moreover, for any given i, the sequence $\{u_{2it}\}_{t\in\mathbb{Z}}$ is i.i.d. over time index t.

[12] This is not true, of course, when modeling extreme events like natural disasters, wars or financial crises. To study these types of series, however, researchers often apply specialized models. Thinking in this direction, a future development could be to extend the framework presented here to allow for innovations with unbounded support.

Assumption 6. (i) $\mathbb{E}[u_{2it}^2|\mathcal{F}_{t-1}]$ is uniformly bounded for all $t \in \mathbb{Z}$ almost surely, (ii) $\mathbb{E}[|u_{2it}|^{2+\delta}] < \infty$ for some $\delta > 0$, (iii) $\mathbb{E}[|Y_{it}|^{2+\delta}]$ is uniformly bounded for all $t \in \mathbb{Z}$ almost surely, and (iv) $\mathbb{E}[Y_{it}^2|\mathcal{F}_{t-1}] < \infty$ for any $\delta > 0$.

Now let $\mathcal{W}_2 \subset \mathbb{R}^d$ be the domain of W_{2t}. By assumption, $\mathcal{W}_2$ is compact and convex and is given by the direct product

$$\mathcal{W}_2 = \mathcal{X}^{1+p} \times \mathcal{Y}^p \times \mathcal{E}_1,$$

where $\mathcal{E}_1$ is the domain of structural innovations ϵ_{1t} i.e. $\mathcal{E} \equiv \mathcal{E}_1 \times \mathcal{E}_2$.

Assumption 7. Define $\zeta_{K,n} := \sup_{w\in\mathcal{W}_2} \|b_\pi^K(w)\|$ and

$$\lambda_{K,n} := [\lambda_{\min}(\mathbb{E}[\, b_\pi^K(W_{2t}) b_\pi^K(W_{2t})' \,])]^{-1/2}.$$

It holds:

(i) There exist $\omega_1, \omega_2 \geqslant 0$ s.t. $\sup_{w\in\mathcal{W}_2} \|\nabla b_\pi^K(w)\| \lesssim n^{\omega_1} K^{\omega_2}$.

(ii) There exist $\overline{\omega}_1 \geqslant 0$, $\overline{\omega}_2 > 0$ s.t. $\zeta_{K,n} \lesssim n^{\overline{\omega}_1} K^{\overline{\omega}_2}$.

(iii) $\lambda_{\min}(\mathbb{E}[\, b^K(W_{2t}) b^K(W_{2t})' \,]) > 0$ for all K and n.

Assumption 7 provides mild regularity conditions on the families of sieves that can be used for the series estimator. More generally, letting $\mathcal{W}_2$ be compact and rectangular makes Assumption 7 hold for commonly used basis functions (Chen and Christensen, 2015).[13] In particular, Assumption 7(i) holds with $\omega_1 = 0$ since the domain is fixed over the sample size.

In the proofs, it is useful to consider the orthonormalized sieve basis. Let

$$\widetilde{b}_\pi^K(w) := \mathbb{E}\left[\, b_\pi^K(W_{2t}) b_\pi^K(W_{2t})' \,\right]^{-1/2} b_\pi^K(w),$$
$$\widetilde{B}_\pi := \left(\widetilde{b}_\pi^K(W_{21}), \ldots, \widetilde{b}_\pi^K(W_{2n})\right)'$$

be the orthonormalized vector of basis functions and the orthonormalized regression matrix, respectively.

Assumption 8. It holds that $\|(\widetilde{B}_\pi' \widetilde{B}_\pi / n) - I_K\| = o_P(1)$.

Assumption 8 is the key assumption imposed by Chen and Christensen (2015) to derive uniform converges rates under dependence. They prove that if $\{W_{2t}\}_{t\in\mathbb{Z}}$ is strictly stationary and β-mixing – with either geometric or algebraic decay, depending on the sieve family of interest – then Assumption 8 holds. Let $(\Omega, \mathcal{Q}, \mathbb{P})$ be the underlying probability space and define

$$\beta(\mathcal{A}, \mathcal{B}) := \frac{1}{2} \sup \sum_{(i,j)\in I\times J} |\mathbb{P}(A_i \cap B_i) - \mathbb{P}(A_i)\mathbb{P}(B_i)|$$

where $\mathcal{A}, \mathcal{B}$ are two σ-algebras, $\{A_i\}_{i\in I} \subset \mathcal{A}$, $\{B_j\}_{j\in J} \subset \mathcal{B}$ and the supremum is taken over all finite partitions of Ω. The h-th β-mixing coefficient of process $\{W_{2t}\}_{t\in\mathbb{Z}}$ is defined as

$$\beta(h) = \sup_t \beta\left(\sigma(\ldots, W_{2t-1}, W_{2t}), , \sigma(W_{2t+h}, W_{2t+h+1}, \ldots)\right),$$

[13] See Chen (2007), Belloni et al. (2015) for additional discussion and examples of sieve families.

and W_{2t} is said to be *geometric* or *exponential* β*-mixing* if $\beta(h) \leqslant \gamma_1 \exp(-\gamma_2 h)$ for some $\gamma_1 > 0$ and $\gamma_2 > 0$. The main issue with mixing assumptions is that they are, in general, hard to compute and evaluate. Therefore, especially in nonlinear systems, assuming that $\beta(h)$ decays exponentially over h imposes very high-level assumptions on the model. There are, however, many setups in which it is known that β-mixing holds under primitive assumptions (see Chen (2013) for examples).

In the next subsection, I will argue that using a different concept of dependence - one rooted in a physical understanding of the underlying stochastic process - leads to imposing transparent assumptions on the model's structure.

3.3 Physical Dependence Conditions

Consider now a *non-structural model* of the form

$$Z_t = G(Z_{t-1}, \epsilon_t). \tag{10}$$

This is a generalization of semi-reduced model (3) where linear and nonlinear components are absorbed into one functional term and B_0 is the identity matrix.[14] Indeed, note that models of the form $Z_t = G(Z_{t-1}, \ldots, Z_{t-p}, \epsilon_t)$ can be rewritten as (10) using a companion formulation. If ϵ_t is stochastic, (10) defines a causal nonlinear stochastic process. More generally, it defines a nonlinear difference equation and an associated dynamical system driven by ϵ_t. Throughout this subsection, I shall assume that $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$ as well as $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$.

Relying on the framework of Pötscher and Prucha (1997), I now introduce explicit conditions that allow to control dependence in nonlinear models by using the toolbox of physical dependence measures developed by Wu (2005, 2011). The aim is to use a dynamical system perspective to address the question of imposing meaningful assumptions on nonlinear dynamic models. This makes it possible to give more primitive conditions under which one can actually estimate (8) in a semi-nonparametric way.

Stability. An important concept for dynamical system theory is that of stability. Stability turns out to play a key role in constructing valid asymptotic theory, as it is well understood in linear models. It is also fundamental in developing the approximation theory of nonlinear stochastic systems.

Example 3.1. (Linear System). As a motivating example, first consider the linear system

$$Z_t = BZ_{t-1} + \epsilon_t$$

where we may assume that $\{\epsilon_t\}_{t\in\mathbb{Z}}$, $\epsilon_t \in \mathbb{R}^{d_Z}$, is a sequence of i.i.d. innovations.[15] It is well-known that this system is stable if and only if the largest eigenvalue of B is strictly less than

[14] In this specific subsection, shock identification does not play a role and, as such, one can safely ignore B_0.

[15] One could alternatively think of the case of a deterministic input, setting $\epsilon_t \sim P_t(a_t)$ where $P_t(a_t)$ is a Dirac density on the deterministic sequence $\{a_t\}_{t\in\mathbb{Z}}$.

one in absolute value (Lütkepohl, 2005). For a higher order linear system, $Z_t = B(L)Z_{t-1} + \epsilon_t$ where $B(L) = B_1 + B_2 L + \ldots + B_p L^{p-1}$, stability holds if and only if $|\lambda_{\max}(\mathbf{B})| < 1$ where

$$\mathbf{B} := \begin{bmatrix} B_1 & B_2 & \cdots & B_p \\ I_{d_Z} & 0 & \cdots & 0 \\ 0 & I_{d_Z} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & I_{d_Z} & 0 \end{bmatrix}$$

is the companion matrix.

Extending the notion of stability from linear to nonlinear systems requires some care. Pötscher and Prucha (1997) derived generic conditions allowing to formally extend stability to nonlinear models by first analyzing *contractive* systems.

Definition 3.1 (Contractive System). *Let $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$, $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$, where $\{Z_t\}_{t\in\mathbb{Z}}$ is generated according to*

$$Z_t = G(Z_{t-1}, \epsilon_t).$$

The system is contractive if for all $(z, z') \in \mathcal{Z} \times \mathcal{Z}$ and $(e, e') \in \mathcal{E} \times \mathcal{E}$

$$\|G(z, \epsilon) - G(z', \epsilon')\| \leqslant C_Z \|z - z'\| + C_\epsilon \|e - e'\|$$

holds with Lipschitz constants $0 \leqslant C_Z < 1$ and $0 \leqslant C_\epsilon < \infty$.

Sufficient conditions to establish contractivity are

$$\sup \left\{ \left\| \text{stack}_{i=1}^{d_Z} \left[\frac{\partial G}{\partial Z}(z^i, e^i) \right]_i \right\| \, \middle| \, z^i \in \mathcal{Z}, e^i \in \mathcal{E} \right\} < 1 \tag{11}$$

and

$$\left\| \frac{\partial G}{\partial \epsilon} \right\| < \infty, \tag{12}$$

where the stacking operator $\text{stack}_{i=1}^{d_Z}[\,\cdot\,]_i$ progressively stacks the rows, indexed by i, of its argument (which can be changing with i) into a matrix. Values $(z^i, e^i) \in \mathcal{Z} \times \mathcal{E}$ change with index i as the above condition is derived using the mean value theorem, therefore it is necessary to consider a different set of values for each component of Z_t.

It is easy to see, as Pötscher and Prucha (1997) point out, that contractivity is often a too strong condition to be imposed. Indeed, even in the simple case of a scalar AR(2) model $Z_t = b_1 Z_{t-1} + b_2 Z_{t-2} + \epsilon_t$, regardless of the values of $b_1, b_2 \in \mathbb{R}$ contractivity is violated. This is due to the fact that in a linear AR(2) model studying contractivity reduces to checking $\|\mathbf{B}\| < 1$ instead of $|\lambda_{\max}(\mathbf{B})| < 1$, and the former is a stronger condition than the latter.[16] One can weaken contractivity – which must hold for G as a map from Z_{t-1} to Z_t – to the idea of *eventual*

[16] See Pötscher and Prucha (1997), pp.68-69.

contractivity. That is, intuitively, one can impose conditions on the dependence of Z_{t+h} on Z_t for $h > 1$ sufficiently large. To do this formally, I first introduce the definition of system map iterates.

Definition 3.2 (System Map Iterates)**.** *Let $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$, $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$ where $\{Z_t\}_{t\in\mathbb{Z}}$ is generated from a sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$ according to*

$$Z_t = G(Z_{t-1}, \epsilon_t).$$

The h-order system map iterate is defined to be

$$\begin{aligned} G^{(h)}(Z_t, \epsilon_{t+1}, \epsilon_{t+2}, \ldots, \epsilon_{t+h}) &:= G(G(\cdots G(Z_t, \epsilon_{t+1}) \cdots, \epsilon_{t+h-1}), \epsilon_{t+h}) \\ &= G(\cdot, \epsilon_{t+h}) \circ G(\cdot, \epsilon_{t+h-1}) \circ \cdots \circ G(Z_t, \epsilon_{t+1}), \end{aligned}$$

where $\circ$ signifies function composition and $G^{(0)}(Z_t) = Z_t$.

To shorten notation, in place of $G^{(h)}(Z_t, \epsilon_{t+1}, \epsilon_{t+2}, \ldots, \epsilon_{t+h})$ I shall use $G^{(h)}(Z_t, \epsilon_{t+1:t+h})$. Additionally, for $1 \leqslant j \leqslant h$, the partial derivative

$$\frac{\partial G^{(h^*)}}{\partial \epsilon_j}$$

for some fixed h^* is to be intended with respect to ϵ_{t+j}, the j-th entry of the input sequence. This derivative does not dependent on the time index since by assumption G is time-invariant and so is $G^{(h)}$.

Taking again the linear autoregressive model as an example,

$$Z_{t+h} = G^{(h)}(Z_t, \epsilon_{t+1:t+h}) = B_1^h Z_t + \sum_{i=0}^{h-1} B_1^i \epsilon_{t+h-i}$$

since $G(z, \epsilon) = B_1 z + \epsilon$. If B_1 determines a stable system, then $\|B_1^h\| \to 0$ as $h \to \infty$ since G^h converges to zero, and therefore $\|B_1^h\| \leqslant C_Z < 1$ for h sufficiently large. It is thus possible to use system map iterates to define stability for higher-order nonlinear systems.

Definition 3.3 (Stable System)**.** *Let $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$, $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$, where $\{Z_t\}_{t\in\mathbb{Z}}$ is generated according to the system*

$$Z_t = G(Z_{t-1}, \epsilon_t).$$

The system is stable if there exists $h^* \geqslant 1$ such that for all $(z, z') \in \mathcal{Z} \times \mathcal{Z}$ and $(e_1, e_2, \ldots e_{h^*}, e'_1, e'_2, \ldots, e'_{h^*}) \in \times_{i=1}^{2h^*} \mathcal{E}$

$$\|G^{(h^*)}(z, e_{1:h^*}) - G^{(h^*)}(z', e'_{1:h^*})\| \leqslant C_Z \|z - z'\| + C_\epsilon \|e_{1:h^*} - e'_{1:h^*}\|$$

holds with Lipschitz constants $0 \leqslant C_Z < 1$ and $0 \leqslant C_\epsilon < \infty$.

It is important to remember that this definition encompasses systems with an arbitrary finite autoregressive structure, i.e., $Z_t = G(Z_{t-p+1}, \ldots, Z_{t-1}, \epsilon_t)$ for $p \geqslant 1$, thanks to the companion formulation of the process. An explicit stability condition, similar to that discussed above for contractivity, can be derived by means of the mean value theorem. Indeed, for a system to be stable it is sufficient that, at iterate h^*,

$$\sup\left\{\left\|\operatorname{stack}_{i=1}^{d_Z}\left[\frac{\partial G^{(h^*)}}{\partial Z}(z^i, e^i_{1:h^*})\right]_i\right\| \,\middle|\, z^i \in \mathcal{Z}, e^i_{1:h^*} \in \bigtimes_{i=1}^{h^*} \mathcal{E}\right\} < 1 \tag{13}$$

and

$$\sup\left\{\left\|\frac{\partial G^{(h^*)}}{\partial \epsilon_j}(z, e_{1:h^*})\right\| \,\middle|\, z \in \mathcal{Z}, e_{1:h^*} \in \bigtimes_{i=1}^{h^*} \mathcal{E}\right\} < \infty, \qquad j = 1, \ldots, h^*. \tag{14}$$

Pötscher and Prucha (1997) have used conditions (11)-(12) and (13)-(14) as basis for uniform laws of large numbers and central limit theorems for L^r-approximable and near epoch dependent processes.

Physical Dependence. Wu (2005) first proposed alternatives to mixing concepts by proposing dependence measures rooted in a dynamical system view of a stochastic process. Much work has been done to use such measures to derive approximation results and estimator properties, see for example Wu et al. (2010), Wu (2011), Chen et al. (2016), and references within.

Definition 3.4. *If for all $t \in \mathbb{Z}$, Z_t has finite rth moment, where $r \geqslant 1$, the functional physical dependence measure Δ_r is defined as*

$$\Delta_r(h) := \sup_t \left\| Z_{t+h} - G^{(h)}(Z'_t, \epsilon_{t+1:t+h}) \right\|_{L^r}$$

where $\|\cdot\|_{L^r} = (\mathbb{E}[\|\cdot\|_r^r])^{1/r}$, Z'_t is due to $\mathcal{F}'_t = (\ldots, \epsilon'_{t-1}, \epsilon'_t)$ and $\{\epsilon'_t\}_{t\in\mathbb{Z}}$ is an independent copy of $\{\epsilon_t\}_{t\in\mathbb{Z}}$.

Chen et al. (2016), among others, show how one may replace the geometric β-mixing assumption with a physical dependence assumption.[17] They show that the key sufficient condition is for $\Delta_r(h)$ to decay sufficiently fast as h grows.

Definition 3.5 (Geometric Moment Contracting Process)**.** *$\{Z_t\}_{t\in\mathbb{Z}}$ is geometric moment contracting (GMC) in L^r norm if there exists $a_1 > 0$, $a_2 > 0$ and $\tau \in (0, 1]$ such that*

$$\Delta_r(h) \leqslant a_1 \exp(-a_2\, h^\tau).$$

GMC conditions can be considered more general than β-mixing, as they encompass well-known counterexamples, e.g., the known counterexample provided by $Z_t = (Z_{t-1} + \epsilon_t)/2$ for ϵ_t i.i.d. Bernoulli r.v.s (Chen et al., 2016). In the following proposition I prove that if contractivity or stability conditions as defined by Pötscher and Prucha (1997) hold for G and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence, then process $\{Z_t\}_{t\in\mathbb{Z}}$ is GMC under weak moment assumptions.

[17] I adapt the definitions of Chen et al. (2016) to work with a system of the form $Z_t = G(Z_{t-1}, \epsilon_t)$.

Proposition 3.1. *Assume that* $\{\epsilon_t\}_{t\in\mathbb{Z}}$, $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$ *are i.i.d. and* $\{Z_t\}_{t\in\mathbb{Z}}$ *is generated according to*

$$Z_t = G(Z_{t-1}, \epsilon_t),$$

where $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$ *and* G *is a measurable function.*

(a) If contractivity conditions (11)-(12) *hold,* $\sup_{t\in\mathbb{Z}}\|\epsilon_t\|_{L^r} < \infty$ *for* $r \geqslant 2$ *and* $\|G(\bar{z}, \bar{\epsilon})\| < \infty$ *for some* $(\bar{z}, \bar{\epsilon}) \in \mathcal{Z} \times \mathcal{E}$*, then* $\{Z_t\}_{t\in\mathbb{Z}}$ *is GMC with*

$$\Delta_r(k) \leqslant a \exp(-\gamma h)$$

where $\gamma = -\log(C_Z)$ *and* $a = 2\|Z_t\|_{L^r} < \infty$.

(b) If stability conditions (13)-(14) *hold,* $\sup_{t\in\mathbb{Z}}\|\epsilon_t\|_{L^r} < \infty$ *for* $r \geqslant 2$ *and* $\|\partial G/\partial Z\| \leqslant M_Z < \infty$*, then* $\{Z_t\}_{t\in\mathbb{Z}}$ *is GMC with*

$$\Delta_r(k) \leqslant \bar{a} \exp(-\gamma_{h*}\, h)$$

where $\gamma_{h*} = -\log(C_Z)/h^*$ *and* $\bar{a} = 2\|Z_t\|_{L^r} \max\{M_Z^{h-1}, 1\}/C_Z < \infty$.

Proposition 3.1 is important in that it links the GMC property to transparent conditions on the structure of the nonlinear model. It also immediately allows handling multivariate systems, while previous work has focused on scalar systems (c.f. Wu (2011) and Chen et al. (2016)).

Finally, it is now possible to show that if $\{W_{2t}\}_{t\in\mathbb{Z}}$ satisfies physical dependence assumptions, then Assumption 8 is fulfilled, c.f. Lemma 2.2 in Chen and Christensen (2015) for β-mixing assumptions.

Lemma 3.1. *If Assumption 7(iii) holds and* $\{W_{2t}\}_{t\in\mathbb{Z}}$ *is strictly stationary and GMC then one may choose an integer sequence* $q = q(n) \leqslant n/2$ *with* $(n/q)^{r+1} q K^\rho \Delta_r(q) = o(1)$ *for* $\rho = 5/2 - (r/2 + 2/r) + \omega_2$ *and* $r > 2$ *such that*

$$\|(\widetilde{B}_\pi' \widetilde{B}_\pi/n) - I_K\| = O_P\left(\zeta_{K,n}\lambda_{K,n}\sqrt{\frac{q\log K}{n}}\right) = o_P(1)$$

provided $\zeta_{K,n}\lambda_{K,n}\sqrt{(q\log K)/n} = o(1)$.

It can be seen that Lemma 3.1 holds by setting $\sqrt{K(\log(n))^2/n} = o(1)$ and choosing $q(n) = \gamma^{-1}\log(K^\rho n^{r+1})$, where γ is the GMC factor introduced in Proposition 3.1. Therefore, the rate is the same as the one derived by Chen and Christensen (2015) for exponentially β-mixing regressors. As shown in Proposition 3.1, system contractivity and stability conditions both imply geometric moment contractivity, meaning that in place of Assumption 8 one may require the following.

Assumption 9. For $r > 2$ it holds either:

(i) $\{Z_t\}_{t\in\mathbb{Z}}$ is GMC in L^r norm,

(ii) $\{Z_t\}_{t\in\mathbb{Z}}$ is generated according to $Z_t = \Phi(Z_{t-1}, \ldots, Z_{t-p}; \epsilon_t)$ where $\sup_{t\in\mathbb{Z}}\|\epsilon_t\|_{L^r} < \infty$ and Φ is either contractive according to Definition 3.1 or stable according to Definition 3.3.

It is straightforward to prove that if GMC conditions are imposed on $\{Z_t\}_{t\in\mathbb{Z}}$, this implies that $\{W_{2t}\}_{t\in\mathbb{Z}}$ is also GMC.[18] Therefore, Lemma 3.1 applies and Assumption 8 as well as Assumption 5(i) are verified.

3.4 Uniform Convergence and Consistency

Since the key asymptotic condition of Chen and Christensen (2015) is upheld under GMC assumptions, their uniform convergence bound on the approximation error of the series estimator can be applied. In order to do so, one must also impose some regularity conditions on π_2.

Without loss of generality, let $\mathcal{X} = [0, 1]$ and let $\|\pi_2\|_\infty := \sup_{w\in\mathcal{Y}}|\pi_2(w)|$ be the sup-norm of the conditional mean function $\pi_2(w)$.

Assumption 10. The unconditional density of X_t is uniformly bounded away from zero and infinity over $\mathcal{X}$.

Assumption 11. For all $1 \leqslant i \leqslant d_Y$ and $0 \leqslant j \leqslant p$, the restriction of g^{21}_{ij} to $[0, 1]$ belongs to the Hölder class $\Lambda^s([0, 1])$ of smoothness $s \geqslant 1$.

Assumptions 10 and 11 are standard in the nonparametric regression literature. One only needs to restrict the complexity of functions g^{21}_{ij} since, for any i, the remainder of $\pi_{2,i}$ consists of linear functions. More precisely, what is really needed is that the nonparametric components of the sieve given by $b_{\pi,1K}, \ldots, b_{\pi,KK}$ are able to approximate g^{21}_{ij} well enough.

Assumption 12. Sieve B_κ belongs to $\mathrm{BSpl}(\kappa, [0, 1]^{d_Y}, r)$, the B-spline sieve of degree r over $[0, 1]^{d_Y}$, or $\mathrm{Wav}(\kappa, [0, 1]^{d_Y}, r)$, the wavelet sieve of regularity r over $[0, 1]^{d_Y}$, with $r > \max\{s, 1\}$.

In the remainder of the paper, I will consider the cubic spline sieve ($r = 3$), but theoretical results are stated in the more general setting. Moreover, d will be the effective dimension of the joint estimation domain for G^{21}_i.

Theorem 3.1 (Chen and Christensen (2015))**.** *Let Assumptions 4, 5, 6, 7, 9, 10, 11 and 12 hold. If*

$$K \asymp (n/\log(n))^{d/(2s+d)},$$

then

$$\|\widehat{\pi}^*_2 - \pi_2\|_\infty = O_P\left((n/\log(n)^{-s/(2s+d)}\right)$$

provided that $\delta \geqslant 2/s$ (in Assumption 6) and $d < 2s$.

In Theorem 3.1 the sup-norm consistency rate generally depends on the dimension d and thus, in principle, the curse of dimensionality slows down convergence compared to parametric estimation. Fortunately, under the current strctural model assumptions, the nonlinear functional components in π_2 are linearly separable in the lag dimension, and thus one may take $d = 1$ as effective dimension. This also means that condition $d < 2s$ is trivially satisfied.

[18] A formal argument can be found in Appendix B.

Two-step Consistency. The following theorem ensures that the two-step estimation procedure produces consistent estimates. Since for impulse response functions one needs to study the iteration of the entire structural model, this results is stated in terms of the full coefficient matrices.

Theorem 3.2. *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be determined by structural model (1). Under Assumptions 1, 4, 5, 6, 7, 9, 10, 11 and 12, let $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$ be the least squares and semi-nonparametric series estimators for Π_1 and Π_2, respectively, based on the two-step procedure. Then,*

$$\|\widehat{\Pi}_1 - \Pi_1\|_\infty = O_P(n^{-1/2})$$

and

$$\|\widehat{\Pi}_2 - \Pi_2\|_\infty \leqslant O_P\left(\zeta_{K,n}\lambda_{K,n}\frac{K}{\sqrt{n}}\right) + \|\widehat{\Pi}_2^* - \Pi_2\|_\infty,$$

where $\widehat{\Pi}_2^*$ is the infeasible series estimator involving ϵ_{1t}.

Sup-norm bounds for $\|\widehat{\Pi}_2^* - \Pi_2\|_\infty$ follow immediately from Lemma 2.3 and Lemma 2.4 in Chen and Christensen (2015). In particular, choosing the optimal nonparametric rate $K \asymp (n/\log(n))^{d/(2s+d)}$ for the infeasible estimator would yield

$$\|\widehat{\Pi}_2^* - \Pi_2\|_\infty = O_P\left((n/\log(n))^{-s/(2s+d)}\right)$$

as per Theorem 3.1. The condition for consistency in Theorem 3.2 reduces to

$$\frac{K^{3/2}}{\sqrt{n}} = o(1),$$

since for B-spline and wavelet sieves $\lambda_{K,n} \lesssim 1$ and $\zeta_{K,n} \lesssim \sqrt{K}$. It simple to show that if for the feasible estimator $\widehat{\Pi}_2$ the same rate $(n/\log(n))^{d/(2s+d)}$ is chosen for K, the consistency condition in the above display is fulfilled assuming $s \geqslant 1$ and $d = 1$.[19]

Remark 3.3. (Hyperparameter Selection). An important practical question when applying any series or kernel-type methods is the selection of hyperparameters. For the former, this entails the choice of the sieve's size K. Although theory provides only asymptotic rates, a number of methods can be used to select K, such as cross-validation, generalized cross-validation and Mallow's criterion (Li and Racine, 2009). In the case of piece-wise splines, once size is selected, knots can be chosen to be the K uniform quantiles of the data. This ensures knots are not located in regions of the domain with very few observations. In simulations and applications, for simplicity, I select sieve sizes manually and locate knots approximately following empirical quantiles. In unreported numerical experiments, I check that results are robust to moderate changes in the number and approximate locations of spline knots.

[19]The rate for K may be optimized by balancing the uniform (infeasible) rate with the error due to residuals. Since this paper is not concerned with finding the optimal rate, I do not perform this exercise here.

4 Impulse Response Analysis

Once the model's linear, functional and structural coefficient are consistently estimated, computation of nonlinear impulse responses must be addressed. As discussed in Section 2, nonlinear IRFs are generally hard to lay hands on, since the functional MA(∞) form of the process is highly non-trivial. In this section, I will provide an explicit, iterative algorithm to compute responses that is numerically straightforward and does not require the construction of moving average functional coefficients. Moreover, since to derive uniform bounds it is assumed that the data has compact support, I will introduce a novel yet familiar IRF definition, called the relaxed impulse response function, which is compatible with boundedness. Lastly, I prove that semi-nonparametric IRF estimates are consistent with respect to their population counterparts.

4.1 Computation

Recall from equation (7) in Section 2.2 that impulse responses involve two moving average lag polynomials, $\Theta(L)$ for the linear model component and $\Gamma(L)$ for the nonlinear component, respectively. As a first step, one can derive a semi-explicit recursive algorithm for computing $\mathrm{IRF}_h(\delta)$ in a manner that does not involve simulations of the innovations process.

Proposition 4.1 (Gonçalves et al. (2021), Proposition 3.1). *Under Assumptions 1, 2 and 3, for any $h = 0, 1, \ldots, H$, let*

$$V_j(\delta) := \mathbb{E}[\Gamma_j X_{t+j}(\delta)] - \mathbb{E}[\Gamma_j X_{t+j}].$$

To compute

$$\mathrm{IRF}_h(\delta) = \Theta_{h,\cdot 1}\delta + \sum_{j=0}^{h} V_j(\delta),$$

the following steps can be used:

(i) For $j = 0$, set $X_t(\delta) = X_t + \delta$ and $V_0(\delta) = \mathbb{E}[\Gamma_0 X_t(\delta)] - \mathbb{E}[\Gamma_0 X_t]$.

(ii) For $j = 1, \ldots, h$, let

$$\begin{aligned} X_{t+j}(\delta) &= X_{t+j} + \Theta_{j,11}\delta + \sum_{k=1}^{j} (\Gamma_{k,11} X_{t+j-k}(\delta) - \Gamma_{j,11} X_{t+j-k}) \\ &= \gamma_j(X_{t+j:t}; \delta), \end{aligned}$$

where γ_j are implicitly defined and depend on $\Theta(L)$ and $\Gamma(L)$.

(iii) For $j = 1, \ldots, h$, compute

$$V_j(\delta) = \mathbb{E}[\Gamma_j \gamma_j(X_{t+j:t}; \delta)] - \mathbb{E}[\Gamma_j X_{t+j}].$$

The proof of Proposition 4.1 is identical to that in Gonçalves et al. (2021), with the only variation being that in the current setup it is not possible to collect the nonlinear function across

$X_{t+j-k}(\delta)$ and X_{t+j-k}. Computation of $X_{t+j}(\delta)$ in step (ii) involves recursive evaluations of nonlinear functions, which is why the algorithm is semi-explicit. For each horizon h, one needs to evaluate $h+1$ iterations of $X_t(\delta)$. Importantly, however, this approach dispenses from the need to simulate innovations $\{\epsilon_{t+j}\}_{j=1}^{h-1}$ as the joint distribution of $\{X_{t+h-1}, X_{t+j-1}, \dots, X_t\}$ already contains all relevant information. Gonçalves et al. (2021) naturally argue that the algorithm outlined in Proposition 4.1 is significantly more efficient than schemes involving Monte Carlo simulations like e.g. the one used by Kilian and Vigfusson (2011).

However, $\{\Gamma_j\}_{j=1}^{h}$ are combinations of real and functional matrices and closed-form derivation is numerically impractical. Note that, by the definition of IRFs, the following *explicit* iterative algorithm is also valid.

Proposition 4.2. *In the same setup of Proposition 4.1, to compute* $\mathrm{IRF}_h(\delta)$ *the following steps can be used:*

(i′) For $j = 0$*, let* $X_t(\delta) = X_t + \delta$ *and*

$$\mathrm{IRF}_0(\delta) = \begin{bmatrix} \delta \\ B_0^{21}\delta \end{bmatrix} + \mathbb{E}\begin{bmatrix} 0 \\ G_{21,0}X_t(\delta) \end{bmatrix} - \mathbb{E}\begin{bmatrix} 0 \\ G_{21,0}X_t \end{bmatrix}.$$

(ii′) For $j = 1, \dots, h$*, let*

$$\begin{aligned} X_{t+j}(\delta) &= \mu_1 + A_{12}(L)Y_{t+j-1}(\delta) + A_{11}(L)X_{t+j-1}(\delta) + \epsilon_{1t+j}, \\ Y_{t+j}(\delta) &= \mu_2 + A_{22}(L)Y_{t+j-1}(\delta) + H_{21}(L)X_{t+j}(\delta) + B_0^{21}\epsilon_{1t+j} + u_{2t+j}, \end{aligned}$$

where $H_{21}(L) := A_{21}(L)L + G_{21}(L)$ *and* $u_{2t+j} := B_0^{22}\epsilon_{2t+j}$*. Setting* $Z_{t+j}(\delta) = (X_t(\delta), Y_t(\delta))'$ *it holds*

$$\mathrm{IRF}_h(\delta) = \mathbb{E}[Z_{t+j}(\delta)] - \mathbb{E}[Z_{t+j}].$$

Proposition 4.2 follows directly from the definition of the unconditional impulse response (6) combined with explicit iteration of the semi-reduced form (2) and sidesteps the MA(∞) formulation in (7). Step (i′) is trivial in nature. Step (ii′) may not seem useful when compared to (ii), since, in practice, innovations ϵ_{1t} and u_{2t} are not available. However, let

$$\widehat{\mu},\ \widehat{A}_{11}(L),\ \widehat{A}_{12}(L),\ \widehat{A}_{21}(L),\ \widehat{H}_{11}(L),\ \widehat{B}_0^{21}$$

be estimates of the model's coefficients derived, for example, from series estimator $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$. In sample, one can compute residuals $\widehat{\epsilon}_{1t}$ and $\widehat{u}_{2t}$, and by definition it holds

$$\begin{aligned} X_t &= \widehat{\mu}_1 + \widehat{A}_{12}(L)Y_{t-1} + \widehat{A}_{11}(L)X_{t-1} + \widehat{\epsilon}_{1t}, \\ Y_t &= \widehat{\mu}_2 + \widehat{A}_{22}(L)Y_{t-1} + \widehat{H}_{21}(L)X_t + \widehat{B}_0^{21}\widehat{\epsilon}_{1t} + \widehat{u}_{2t}. \end{aligned}$$

This means that one can readily construct the shocked sequence recursively as

$$\begin{aligned} \widehat{X}_{t+j}(\delta) &= \widehat{\mu}_1 + \widehat{A}_{12}(L)\widehat{Y}_{t+j-1}(\delta) + \widehat{A}_{11}(L)\widehat{X}_{t+j-1}(\delta) + \widehat{\epsilon}_{1t+j}, \\ \widehat{Y}_{t+j}(\delta) &= \widehat{\mu}_2 + \widehat{A}_{22}(L)\widehat{Y}_{t+j-1}(\delta) + \widehat{H}_{21}(L)\widehat{X}_{t+j}(\delta) + \widehat{B}_0^{21}\widehat{\epsilon}_{1t+j} + \widehat{u}_{2t+j}, \end{aligned}$$

for $j = 1, \ldots, h$ where $\widehat{X}_t(\delta) = X_t + \delta$, $\widehat{X}_{t-s} = X_{t-s}$ for all $s \geqslant 1$ and similarly for $\widehat{Y}_t(\delta)$. To evaluate a structural IRF, over a sample of size n one can compute

$$\widehat{\mathrm{IRF}}_h(\delta) = \frac{1}{n-j} \sum_{t=1}^{n-j} \left[\widehat{Y}_{t+j}(\delta) - Y_t \right],$$

which is still considerably less demanding than Monte Carlo simulations. Additionally, the advantage in implementing steps (i′)-(ii′) over the procedure in Proposition 4.1 is that, when $\widehat{H}_{21}(L)$ is a semi-nonparametric estimate, iterating model equations is numerically much more straightforward than handling functional MA matrices $\{\widehat{\Gamma}_j\}_{j=1}^h$.

4.2 Nonlinear Responses with Relaxed Shocks

Following Proposition 4.1, the sample impulse response would be

$$\widehat{\mathrm{IRF}}_h(\delta) := \widehat{\Theta}_{h,\cdot 1}\delta + \sum_{j=0}^{h} \bar{V}_j(\delta), \tag{15}$$

where

$$\bar{V}_j(\delta) := \frac{1}{n-j} \sum_{t=1}^{n-j} \left[\widehat{\Gamma}_j \widehat{\gamma}_j(X_{t+j:t}; \delta) - \widehat{\Gamma}_j X_{t+j} \right]$$

and $\widehat{\Theta}$, $\widehat{\Gamma}$ and $\widehat{\gamma}_j$ are plug-in estimates of the respective quantities based on $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$. However, under Assumptions 4 and 5, the construction of impulse response (15) is improper. This can be immediately seen by noticing that, at impact,

$$X_t(\delta) = \gamma_j(X_t; \delta) = X_t + \delta,$$

meaning that $\mathbb{P}(X_t(\delta) \notin \mathcal{X}) > 0$ since there is a translation of size δ in the support of X_t. The problem is rooted in the fact that the standard definition of IRF involves a translation of the distribution of time t structural innovations, which is incompatible with the assumptions imposed in Section 3 to derive semi-nonparametric consistency.

There are multiple ways to address this issue. One option, which would require substantial technical work, is to extend Theorem 3.2 to encompass regressors with unbounded or expanding domains. A potential direction could be coupling the weighted sieves of Chen and Christensen (2015) with appropriately defined shocks. Instead, I propose to take a more direct approach by changing the *type* of structural shock one studies in a way consistent with bounded domains for all variables.

Definition 4.1. *A mean-shift structural shock $\epsilon_{1t}(\delta)$ is a transformation of ϵ_{1t} such that*

$$\mathbb{P}(\epsilon_{1t}(\delta) \in \mathcal{E}_1) = 1 \quad \text{and} \quad \mathbb{E}[\epsilon_{1t}(\delta)] = \delta.$$

A mean-shift shock is such that the distribution of time t innovations is shifted to have mean δ, while retaining support $\mathcal{E}$ almost surely. This definition is natural in that it makes evaluating the effect of the MA(∞) component of the unconditional IRF straightforward. With

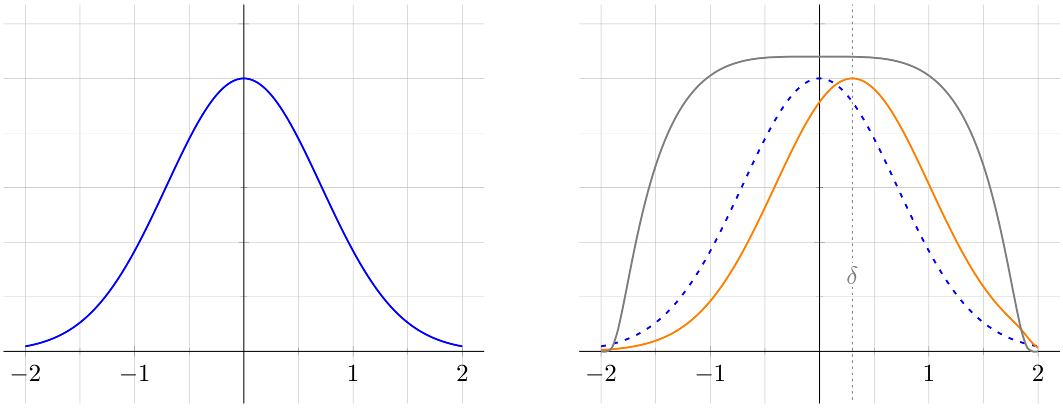


Figure 1: Example of symmetric shock relaxation. Unperturbed (left, blue) versus shocked (right, orange) densities of innovations ϵ_{1t}. The shock relaxation function (right, gray) and δ together determine the form of the relaxed shock used to compute the IRF.

a mean-shift shock, at impact it holds

$$X_t(\delta) = X_t + \epsilon_{1t}(\delta) - \epsilon_{1t},$$

yet $\epsilon_{1t}(\delta) - \epsilon_{1t}$ is not known unless the transformation for the mean-shock is itself known. Unfortunately, the assumption that the mean of $\epsilon_{1t}(\delta)$ is exactly equal to δ requires that the distribution of ϵ_{1t} be known to properly choose a mean-shift transform. If instead one is willing to assume only that $\mathbb{E}[\epsilon_{1t}(\delta)] \approx \delta$, it is possible to sidestep this requirement by introducing a *shock relaxation function.*

Definition 4.2 (Shock Relaxation Function)**.** *A shock relaxation function is a map* $\rho : \mathcal{E}_1 \to [0, 1]$ *such that* $\rho(z) = 0$ *for all* $z \in \mathbb{R} \setminus \mathcal{E}_1$, $\rho(z) \geqslant 0$ *for all* $z \in \mathcal{E}_1$ *and there exists* $z_0 \in \mathcal{E}_1$ *for which* $\rho(z_0) = 1$.

In general, choosing a shock relaxation function without taking into account the shape of domain $\mathcal{E}_1$ does not necessarily imply that the relaxed shocks will not push the structural variable out-of-bounds. Therefore, I also introduce the notion of *compatibility.*

Definition 4.3 (Compatible Relaxation)**.** *Consider a shock* $\delta \in \mathbb{R}$ *and let* $\mathcal{E}_1 = [a, b]$.

(i) If $\delta > 0$, ρ *is said to be right-compatible with* δ *if*

$$\rho(z) \leqslant \frac{b - z}{|\delta|} \text{for all } z \in \mathcal{E}.$$

(ii) If $\delta < 0$, ρ *is said to be left-compatible with* δ *if*

$$\rho(z) \leqslant \frac{a + z}{|\delta|} \quad \text{for all } z \in \mathcal{E}.$$

(iii) Given shock size $|\delta| > 0$, ρ *is said to be compatible if it is both right- and left-compatible.*

By setting

$$\epsilon_{1t}(\delta) = \epsilon_{1t} + \delta\rho(\epsilon_{1t})$$

where ρ is compatible with δ, it follows that $X_t(\delta) = X_t + \delta\rho(\epsilon_{1t})$ and $|\mathbb{E}[\epsilon_{1t}(\delta)]| = |\delta\mathbb{E}[\rho(\epsilon_{1t})]| \leqslant |\delta|$ since $\mathbb{E}[\rho(\epsilon_{1t})] \in [0, 1)$ by definition of ρ. If ρ is a bump function, a relaxed shock is a structural shock that has been mitigated proportionally to the density of innovations at the edges of $\mathcal{E}_1$ and the squareness of ρ. For better intuition, Figure 1 provides a graphical rendition of shock relaxation of a symmetric error distribution with a bump function.

Remark 4.1. The definition of compatible relaxation function is *static*, as it considers only the impact effect of a shock. Nonetheless, the assumption that $X_t \in \mathcal{X}$ for all t must also hold for $X_t(\delta)$, the shocked structural variable. In theory, given δ, one can always either expand $\mathcal{X}$ or strengthen ρ so that compatibility is enforced at all horizons $1 \leqslant h \leqslant H$. For simulations, where one has access to the data generating process, the choice of domains and relaxation functions can be done transparently. In practice, some care is required. When working with empirical data, unless one is willing to assume X_t is wholly exogenous – as in Section 6.1 with monetary policy shocks – or strictly autoregressive, some scenarios are more amenable to analysis with the framework presented here than other. In Section 6.2, following Istrefi and Mouabbi (2018), I will let X_t be a non-negative uncertainty measure, so that negative shocks are harder to study without producing sequences that contain *negative* uncertainty values. Thus, I will focus on positive, contractionary shocks.

For a given X_t, transformation $X_t+\delta\rho(\epsilon_{1t})$ is not directly applicable since ϵ_{1t} is not observed. In practice, therefore, I will consider

$$\widehat{X}_t(\delta) := X_t + \delta\rho(\widehat{\epsilon}_{1t}).$$

For simplicity of notation, let $\widetilde{\delta}_t := \delta\rho(\epsilon_{1t})$. Similarly to Step (ii) of Proposition 4.1, given a path $X_{t+j:t}$ one finds

$$\begin{aligned} X_{t+j}(\widetilde{\delta}_t) &= X_{t+j} + \Theta_{j,11}\widetilde{\delta}_t + \sum_{k=1}^{j} (\Gamma_{k,11}X_{t+j-k}(\widetilde{\delta}_t) - \Gamma_{k,11}X_{t+j-k}) \\ &= \gamma_j(X_{t+j:t}; \widetilde{\delta}_t), \end{aligned}$$

The relaxed-shock impulse response is thus given by

$$\widetilde{\mathrm{IRF}}_h(\delta) := \mathbb{E}[Z_{t+j}(\widetilde{\delta}_t) - Z_{t+j}] = \Theta_{h,\cdot 1}\delta\, \mathbb{E}\left[\rho(\epsilon_{1t})\right] + \sum_{k=1}^{j} \mathbb{E}\left[\Gamma_k X_{t+j-k}(\widetilde{\delta}_t) - \Gamma_k X_{t+j-k}\right].$$

In what follows, I show that by replacing $\widetilde{\delta}_t$ with $\widehat{\widetilde{\delta}}_t = \delta\rho(\widehat{\epsilon}_{1t})$ it is possible to consistently estimate unconditional expectations involving $X_{t+j}(\widetilde{\delta}_t)$ as well as X_{t+j}, and thus $\widetilde{\mathrm{IRF}}_h(\delta)$, by averaging over sample realizations.

4.3 Relaxed Impulse Response Consistency

For a given $\delta \in \mathbb{R}$ and compatible shock relaxation function ρ, vector $V_j(\delta)$ is the nonlinear component of impulse responses. One can focus on a specific variable's response by introducing, for $1 \leqslant \ell \leqslant d$,

$$V_{j,\ell}(\delta) := \frac{1}{n-j} \sum_{t=1}^{n-j} \left[\Gamma_{j,\ell} \gamma_j(X_{t+j:t}; \widetilde{\delta}_t) - \Gamma_{j,\ell} X_{t+j} \right],$$

where $V_{j,\ell}(\delta)$ is the horizon j nonlinear effect on the ℓth variable and $\Gamma_{j,\ell}$ is the ℓth component of functional vector Γ_j. For the sake of notation I also define

$$v_{j,\ell}(X_{t+j:t}; \widetilde{\delta}_t) := \Gamma_{j,\ell} \gamma_j(X_{t+j:t}; \widetilde{\delta}_t) - \Gamma_{j,\ell} X_{t+j}.$$

Let $\widehat{v}_{j,\ell}(X_{t+j:t}; \widehat{\widetilde{\delta}}_t)$ be its sample equivalent, so that

$$\widehat{v}_{j,\ell}(X_{t+j:t}; \widehat{\widetilde{\delta}}_t) = \widehat{\Gamma}_{j,\ell} \widehat{\gamma}_j(X_{t+j:t}; \widehat{\widetilde{\delta}}_t) - \widehat{\Gamma}_{j,\ell} X_{t+j},$$

$$\widehat{V}_{j,\ell}(\delta) = \frac{1}{n-j} \sum_{t=1}^{n-j} \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widehat{\widetilde{\delta}}_t\big)$$

and

$$\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) = \Theta_{h,\cdot 1} \delta \, n^{-1} \sum_{t=1}^{n} \rho(\widehat{\epsilon}_{1t}) + \sum_{j=0}^{h} \widehat{V}_{j,\ell}(\delta)$$

for $1 \leqslant \ell \leqslant d$.

Theorem 4.1. *Let $\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta)$ be a semi-nonparametric estimate for the horizon h relaxed shock IRF of variable ℓ. Under the same assumptions as in Theorem 3.2*

$$\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) \xrightarrow{P} \widetilde{\mathrm{IRF}}_{h,\ell}(\delta)$$

for any fixed integers $0 \leqslant h < \infty$ and $1 \leqslant \ell \leqslant d$.

5 Simulations

This section is devoted to analyzing the empirical performance of the two-step semi-nonparametric estimation strategy discussed above. I will consider the two simulation setups employed by Gonçalves et al. (2021), with focus on bias and MSE of the estimated relaxed shocked impulse response functions. Additionally, I provide simulations under a modified design which highlight how in larger samples the non-parametric sieve estimator consistently recovers impulse responses, while a least-squares estimator constructed with a pre-specified nonlinear transform does not. In all simulations, I use a B-spline sieve of order 1.

5.1 Benchmark Bivariate Design

The first simulation setup involves a bivariate DGP where the structural shock does not directly affect other observables. This is a simple environment to check that indeed the two-step estimator recover the nonlinear component of the model and impulse responses are consistently estimated, and that the MSE does not worsen excessively.

I consider three bivariate data generation processes. DGP 1 sets X_t to be a fully exogenous innovation process,

$$\begin{aligned}
X_t &= \epsilon_{1t}, \\
Y_t &= 0.5Y_{t-1} + 0.5X_t + 0.3X_{t-1} - 0.4\max(0, X_t) + 0.3\max(0, X_{t-1}) + \epsilon_{2t}.
\end{aligned} \tag{16}$$

DGP 2 adds an autoregressive component to X_t, but maintains exogeneity,

$$\begin{aligned}
X_t &= 0.5X_{t-1} + \epsilon_{1t}, \\
Y_t &= 0.5Y_{t-1} + 0.5X_t + 0.3X_{t-1} - 0.4\max(0, X_t) + 0.3\max(0, X_{t-1}) + \epsilon_{2t}.
\end{aligned} \tag{17}$$

Finally, DGP 3 add an endogenous effect of Y_{t-1} on the structural variable by setting

$$\begin{aligned}
X_t &= 0.5X_{t-1} + 0.2Y_{t-1} + \epsilon_{1t}, \\
Y_t &= 0.5Y_{t-1} + 0.5X_t + 0.3X_{t-1} - 0.4\max(0, X_t) + 0.3\max(0, X_{t-1}) + \epsilon_{2t}.
\end{aligned} \tag{18}$$

Following Assumption 1, innovations are mutually independent. To accommodate Assumptions 4 and 5, both ϵ_{1t} and ϵ_{2t} are drawn from a truncated standard Gaussian distribution over $[-3, 3]$.[20] All DGPs are centered to have zero intercept in population.

I evaluate bias and MSE plots using 1000 Monte Carlo simulation. For a chosen horizon H, the impact of a relaxed shock on ϵ_{1t} is evaluated on Y_{t+h} for $h = 1, \ldots, H$. To compute the population IRF, I employ a direct simulation strategy that replicates the shock's propagation through the model and I use 10 000 replications. To evaluate the estimated IRF, the two-step procedure is implemented: a sample of length n is drawn, the linear least squares and the semi-nonparametric series estimators of the model are used to estimate the model and the relaxed IRF is computed following Proposition 4.2. For the sake of brevity, I discuss the case of $\delta = 1$ and I set the shock relaxation function to be

$$\rho(z) = \exp\left(1 + \left[\left|\frac{z}{3}\right|^4 - 1\right]^{-1}\right)$$

over interval $[-3, 3]$ and zero everywhere else.[21] Choices of $\delta = -1$ and $\delta = \pm 0.5$ yield similar results in simulations, so I do not discuss them here.

Figure 2 contains the results for sample size $n = 240$. This choice is motivated by considering the average sample sizes found in most macroeconometric settings: it is equivalent to 20 years of monthly data or 60 yearly of quarterly data (Gonçalves et al., 2021). The benchmark method is an OLS regression that relies on a priori knowledge of the underlying DGP specification. Given the moderate sample size, to construct the cubic spline sieve estimator of the nonlinear component of the model I use a single knot, located at 0. The simulations in Figure 2 show that while the MSE is slighlty higher for the sieve model, the bias is comparable across methods. Note

[20] Let $e_{it} \sim \mathcal{N}(0, 1)$ for $i = 1, 2$, then the truncated Gaussian innovations used in simulation are set to be $\epsilon_{it} = \min(\max(-3, e_{it}), 3)$. The resulting r.v.s have a non-continuous density with two mass points at -3 and 3. However, in practice, since these masses are negligible, for the moderate sample sizes used this choice does not create issues.

[21] It can be easily checked that this choice of ρ is compatible with shocks of size $0 \leqslant |\delta| \leqslant 1$.

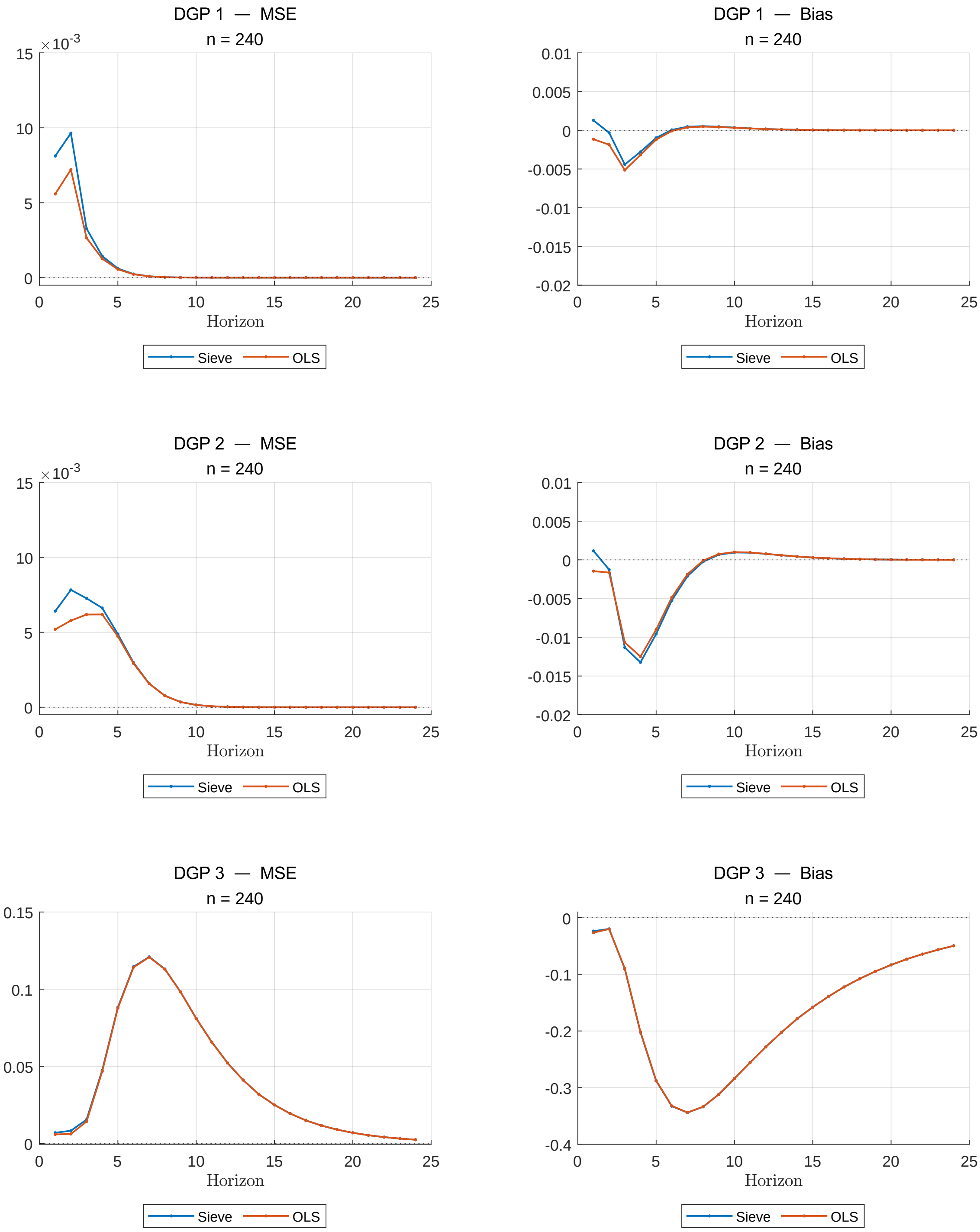


Figure 2: Simulations results for DGPs 1-3.

that for DGP 3, due to the dependence of the structural variable on non-structural series lags, the MSE and bias increase significantly, and there is no meaningful difference in performance between the two estimation approaches.

5.2 Structural Partial Identification Design

To showcase the validity of the proposed sieve estimator under the type of partial structural identification discussed in the paper, I again rely on the simulation design proposed by Gonçalves et al. (2021). All specifications are block-recursive, and require estimating the contemporaneous effects of a structural shock on non-structural variables, unlike in the previous section.

The form of the DGPs is

$$B_0 Z_t = B_1 Z_{t-1} + C_0 f(X_t) + C_1 f(X_{t-1}) + \epsilon_t,$$

where in all variations of the model

$$B_0 = \begin{bmatrix} 1 & 0 & 0 \\ -0.45 & 1 & -0.3 \\ -0.05 & 0.1 & 1 \end{bmatrix}, \quad C_0 = \begin{bmatrix} 0 \\ -0.2 \\ 0.08 \end{bmatrix}, \quad \text{and} \quad C_1 = \begin{bmatrix} 0 \\ -0.1 \\ 0.2 \end{bmatrix}.$$

I focus on the case $f(x) = \max(0, x)$, since this type of nonlinearity is simpler to study. DGP 4 treats X_t as an exogenous shock by setting

$$B_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0.15 & 0.17 & -0.18 \\ -0.08 & 0.03 & 0.6 \end{bmatrix};$$

DGP 5 add serial correlation to X_t,

$$B_1 = \begin{bmatrix} -0.13 & 0 & 0 \\ 0.15 & 0.17 & -0.18 \\ -0.08 & 0.03 & 0.6 \end{bmatrix};$$

and DGP 6 includes dependence on Y_{t-1},

$$B_1 = \begin{bmatrix} -0.13 & 0.05 & -0.01 \\ 0.15 & 0.17 & -0.18 \\ -0.08 & 0.03 & 0.6 \end{bmatrix}.$$

For these data generating processes, I employ the same setup of simulations with DGPs 1-3, including the number of replications as well as the type of relaxed shock. as well as the sieve grid. Here too I evaluate MSE and bias of both the sieve and the correct specification OLS estimators with as sample size of $n = 240$ observations. The results in Figure 3 show again that there is little difference in terms of performance between the semi-nonparametric sieve approach and a correctly-specified OLS regression.

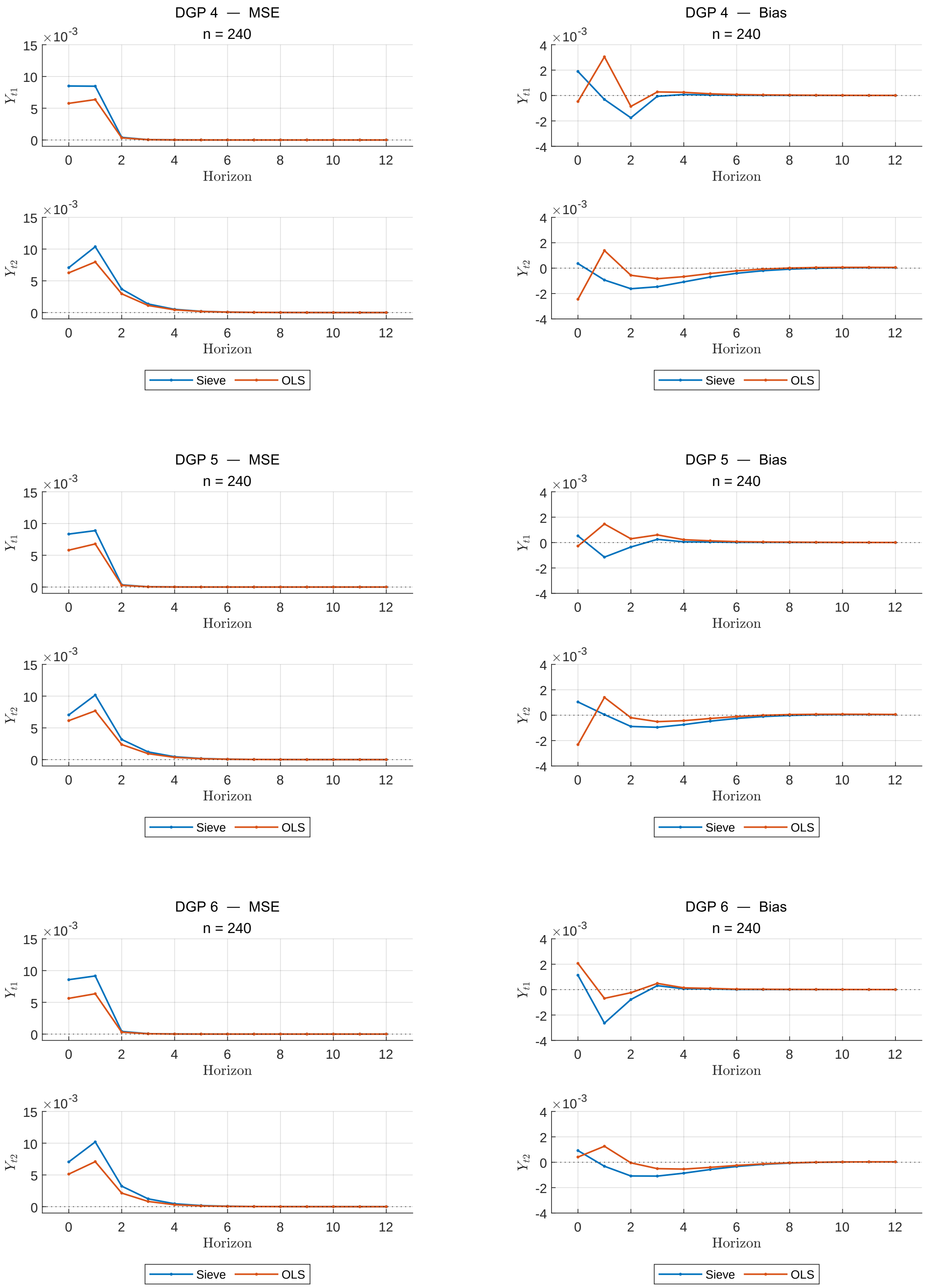


Figure 3: Simulations results for DGPs 4-6.

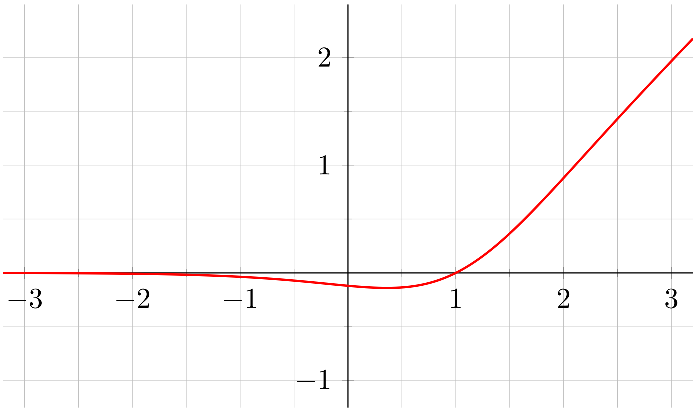


Figure 4: Plot of nonlinear function $\varphi(x)$ used in DGP 7.

5.3 Model Misspecification

The previous sections report results that support the use of the sieve IRF estimator in a sample of moderate size, since it performs comparably to a regression performed with a priori knowledge of the underlying DGP. I now show that the semi-nonparametric approach is also robust to model misspecification compared to simpler specifications involving fixed choices for nonlinear transformations.

To this end, I modify DGP 2 to use a smooth nonlinear transformation to define the effect of structural variable X_t on Y_t. That is, there is no compounding of linear and nonlinear effects. The autoregressive coefficient in the equation for X_t is also increased to make the shock more persistent. The new data generating process, DGP 2′, is, thus, given by

$$\begin{aligned} X_t &= 0.8X_{t-1} + \epsilon_{1t}, \\ Y_t &= 0.5Y_{t-1} + 0.9\varphi(X_t) + 0.5\varphi(X_{t-1}) + \epsilon_{2t}. \end{aligned} \tag{19}$$

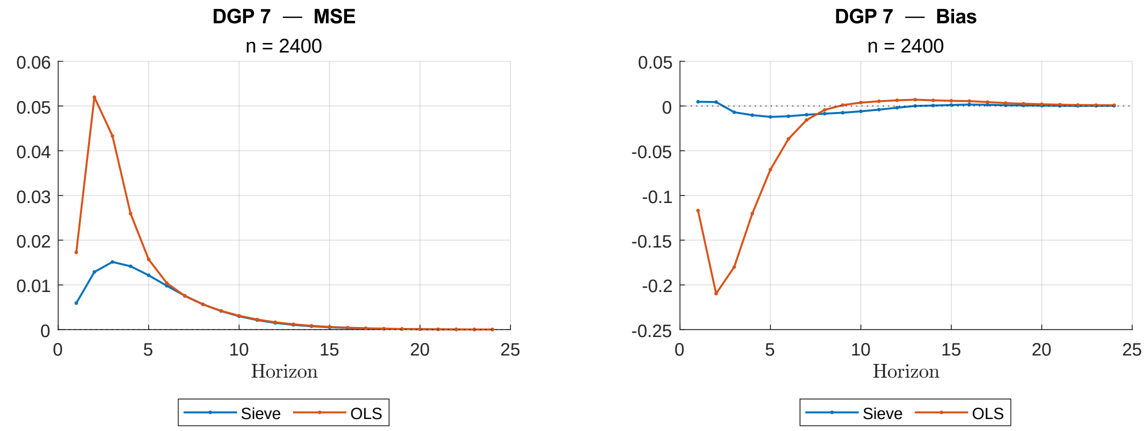
where $\varphi(x) := (x-1)(0.5 + \tanh(x-1)/2)$, which is plotted in Figure 4.

To emphasize the difference in estimated IRFs, in this setup I focus on $\delta = \pm 2$, which requires adapting the choice of innovations and shock relaxation function. In simulations of DGP 2′, ϵ_{1t} and ϵ_{2t} are both drawn from a truncated standard Gaussian distribution over $[-5, 5]$. The shock relaxation function of this setup is given by

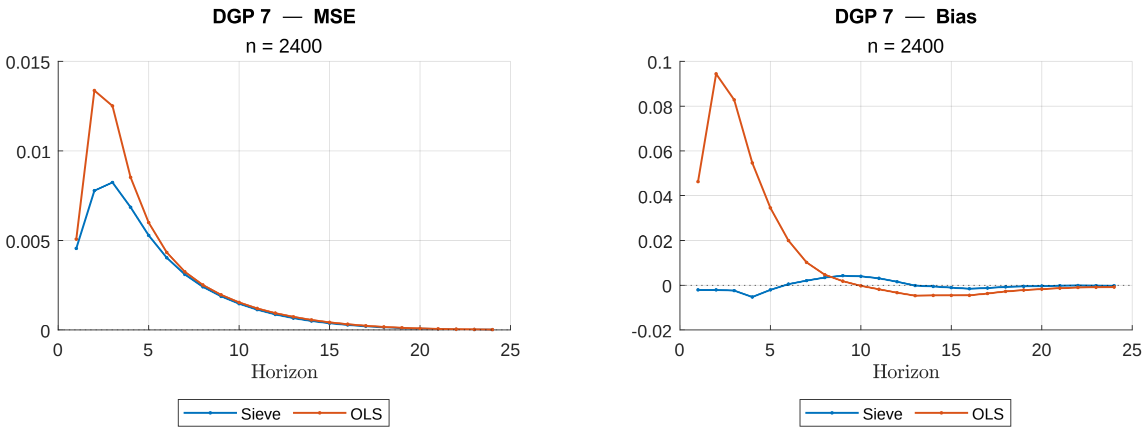
$$\rho(z) = \exp\left(1 + \left[\left|\frac{z}{5}\right|^{3.9} - 1\right]^{-1}\right)$$

over interval $[-5, 5]$ and zero everywhere else. This form of ρ is adapted to choices of δ such that $0 < |\delta| \leqslant 2$. The sieve grid now consists of 4 equidistant knots within $(-5, 5)$. I use the same numbers of replications as in the previous simulations. Finally, the regression design is identical to that used for DGP 2 under correct specification.

The results obtained with sample size $n = 2400$ are collected in Figure 5. I choose this larger sample size to clearly showcase the inconsistency of impulse responses under misspecification: as it can be observed, the simple OLS estimator involving the negative-censoring transform



(a) $\delta = +2$



(b) $\delta = -2$

Figure 5: Simulations results for DGP 7.

produces IRF estimates with consistently worse MSE and bias than those of the sieve estimator at almost all horizons. Similar results are also obtained for more moderate shocks $\delta = \pm 1$, but the differences are less pronounced. These simulations suggest that the semi-nonparametric sieve estimator can produce substantially better IRF estimates in large samples than methods involving nonlinear transformations selected a priori.

In this setup, it is also important to highlight the fact that the poor performance of OLS IRF estimates does not come from $\varphi(x)$ being "complex", and, thus, hard to approximate by combinations of simple functions. In fact, if in DGP $2'$ function φ is replaced by $\widetilde{\varphi}(x) := \varphi(x+1)$, the differences between sieve and OLS impulse response estimates become minimal in simulations, with the bias of the latter decreasing by approximately an order of magnitude (see Figure 8 in Appendix C). This is simply due the fact that $\widetilde{\varphi}(x)$ is well approximated by $\max(0, x)$ directly. However, one then requires either prior knowledge or sheer luck when constructing the nonlinear transforms of X_t for an OLS regression. The proposed series estimator, instead, just requires an appropriate choice of sieve. Many data-driven procedures to select sieves in applications have been proposed, see for example the discussion in Kang (2021).

6 Empirical Applications

In this section, I showcase the practical utility of the proposed semi-nonparametric sieve estimator by considering two applied exercises. First, I revisit the empirical analysis of Gonçalves et al. (2021), which is itself based on the work of Tenreyro and Thwaites (2016). This provides both linear and nonlinear benchmarks for the monetary policy responses within a compact econometric model. I find that, although the differences between approaches are mild, nonparametric IRFs in fact provide counter-evidence to the conclusions reported by Gonçalves et al. (2021). In the second application, I compare the linear and nonlinear impulse responses that are produced by uncertainty shocks in the setup studied by Istrefi and Mouabbi (2018). Here, sieve-estimated IRFs show differences in shape, timing and intensity, chiefly when the sign of the shock changes.

6.1 Monetary Policy Shocks

The objective of the empirical analysis in Gonçalves et al. (2021) is to analyze the effects of a monetary policy shock on a model of the US macroeconomy. Structural identification is achieved via a narrative approach, following the seminal work of Romer and Romer (2004).

The four-variable model is set up identically to the one of Gonçalves et al. (2021), Section 7. Let $Z_t = (X_t, \mathrm{FFR}_t, \mathrm{GDP}_t, \mathrm{PCE}_t)'$, where X_t is the series of narrative U.S. monetary policy shocks, FFR_t is the federal funds rate, GDP_t is log real GDP and PCE_t is PCE inflation.[22] As

[22] In Gonçalves et al. (2021) p. 122, it is mentioned that CPI inflation is included in the model, but both in the replication package made available by one the authors (`https://sites.google.com/site/lkilian2019/research/code`) from which I source the data, and Tenreyro and Thwaites (2016), PCE inflation is used instead. Moreover, the authors say that both the FFR and PCE enter the model in first differences, yet in their code these variables are kept in levels. I keep their original formulation to allow for a proper comparison between estimation methods.

a pre-processing step, GDP is transformed to log GDP and then linearly detrended. The data is available quarterly and spans from 1969:Q1 to 2007:Q4. As in Tenreyro and Thwaites (2016), I use a model with one lag, $p = 1$. Narrative shock X_t is considered to be an i.i.d. sequence, i.e. $X_t = \epsilon_{1t}$, therefore I assume no dependence on lagged variables when implementing pseudo-reduced form (2). Like in Gonçalves et al. (2021), I consider positive and negative shocks of size $|\delta| = 1$. As such, I choose

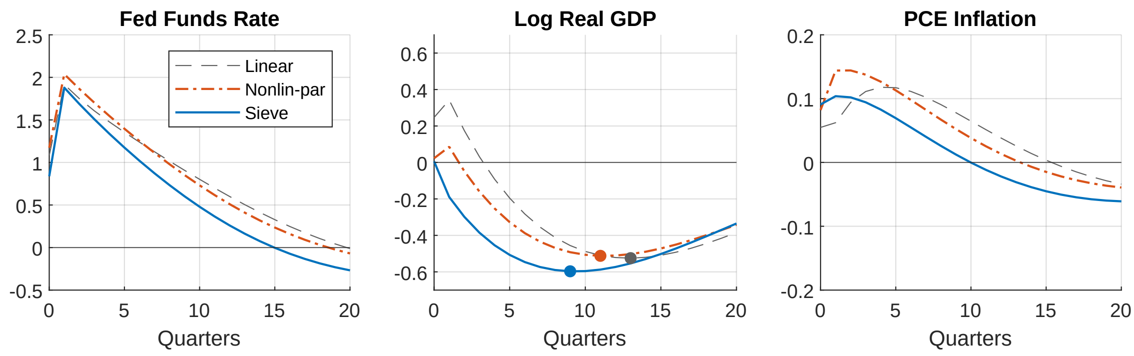
$$\rho(z) = \mathbb{I}\{|z| \leqslant 4\} \exp\left(1 + \left[\left|\frac{z}{4}\right|^6 - 1\right]^{-1}\right)$$

to be the shock relaxation functions. Figure 10 in Appendix C provides a check for the validity of ρ given the sample distribution of X_t. Knots for sieve estimation are located at $\{-1, 0, 1\}$. The model is block-recursive, and the structural formulation of Section 2.2 allows identifying the U.S. monetary policy shocks without the need to impose additional assumptions on the remaining shocks. Gonçalves et al. (2021), following Tenreyro and Thwaites (2016), use two nonlinear transformations, $F(x) = \max(0, x)$ and $F(x) = x^3$, to try to gauge how negative versus positive and large versus small shocks, respectively, affect the U.S. macroeconomy. For clarity, below I refer to this approach as "parametric nonlinear method". Since the authors find that the inclusion of a cubic term does not meaningfully change impulse responses, I focus on comparing the IRFs estimated via sieve regression with the ones obtained by setting $F(x) = \max(0, x)$, as well as by not including nonlinear terms (i.e. linear IRFs).

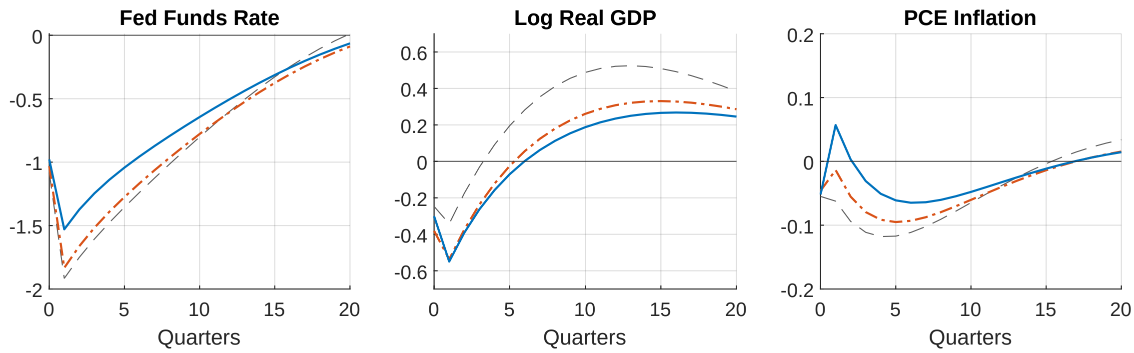
Figure 6 shows the estimated impulse response to both a positive and negative unforeseen monetary policy shock. The impact on the federal funds rate is consistent across all three procedures, but there are important differences in GDP and inflation responses. In case of an exogenous monetary tightening change, the parametric nonlinear response for GDP, unlike in the case of linear and parametric nonlinear IRFs, is nearly zero at impact and has a monotonic decrease until around 10 quarters ahead. The change is shape is meaningful, as the procedure of Gonçalves et al. (2021) still yields a small short-term upward jump in GDP when a monetary tightening shock hits. Moreover, after the positive shock, the sieve GDP responses reaches its lowest value 4 and 2 quarters before the linear and parametric nonlinear responses, while its size is 13% and 16% larger, respectively.[23] Finally, the sieve PCE response is positive for a shorter interval, but looks to be more persistent once it turns negative also 10 months after impact.

When the shock is expansionary, sieve IRFs show a pronounced asymmetry, even more than that of parametric nonlinear responses. One can observe that semi-nonparametric federal funds rate IRF is marginally mitigated compared to the alternative estimates. An important puzzle is due to the clearly negative impact on GDP. Indeed, both types of nonlinear responses show a drop in output in the first 5 quarters. Also note that the PCE inflation has a positive spike the first couple of quarters after impact. Such a quick change seems unrealistic, as one does not expect inflation to suddenly reverse sign, but, as Gonçalves et al. (2021) also remark, the overall

[23]The strength of this effect changes across different shocks sizes, as Figure 12 in Appendix C proves. As shocks sizes get smaller, nonlinear IRFs, both parametric and sieve, show decreasing negative effects.



(a) $\delta = +1$



(b) $\delta = -1$

Figure 6: Effect of an unexpected U.S. monetary policy shock on federal funds rate, GDP and inflation. Linear (gray, dashed), parametric nonlinear with $F(x) = \max(0, x)$ (red, point-dashed) and sieve (blue, solid) structural impulse responses. For $\delta = +1$, the lowest point of the GDP response is marked with a dot.

impact on inflation of both shocks is small when compared to the change in federal funds rate.

This comparison between methods, and specifically the nature of nonparametric impulse responses, provides evidence that a small econometric model, such as the one studied by Tenreyro and Thwaites (2016), may be inadequate to fully capture the dynamic effects of monetary policy shocks. In both setups, however, impulse response interpretation is only suggestive, as confidence bands are missing and only pointwise IRFs are available. Whether the puzzles highlighted above would persist after accounting for estimation uncertainty is an important research question that I leave for future analysis.

6.2 Uncertainty Shocks

Uncertainty in interest rates appears to be a significant factor in recent economic history. Starting with the fundamental changes brought forth by the unprecedented measures of unconventional monetary policy after the 2007-2008 financial crisis, to the powerful economic stimuli during the COVID-19 pandemic, and finally the subsequent interest rate tightening and inflation phenomenon of 2022, central banks and institutional agents are often very concerned about uncertainty. Since traditional central bank policymaking is heavily guided by the principle that the central bank *can* and *should* influence expectations, controlling the (perceived) level of ambiguity in current and future commitments is key.

Istrefi and Mouabbi (2018) provide an analysis of the impact of unforeseen changes in the level of subjective interest rate uncertainty on the macroeconomy. They derive a collection of new indices based on short- and long-term profession forecasts. Their empirical study goes in depth into studying the different components that play a role in transmitting uncertainty shocks, but here I will focus on re-evaluating their structural impulse response estimates under the light of potentially-missing nonlinear effects. For the sake of simplicity, my evaluation will focus only on the 3-months-ahead uncertainty measure for short-term interest rate maturities (3M3M) and the US economy.[24]

Like in Istrefi and Mouabbi (2018), let $Z_t = (X_t, \mathrm{IP}_t, \mathrm{CPI}_t, \mathrm{PPI}_t, \mathrm{RT}_t, \mathrm{UR}_t)'$ be a vector where X_t is the chosen uncertainty measure, IP_t is the (log) industrial production index, CPI_t is the CPI inflation rate, PPI_t is the producer price inflation rate, RT_t is (log) retail sales and UR_t is the unemployment rate. The nonlinear model specification is given by

$$Z_t = \mu + A_1 Z_{t-1} + A_2 Z_{t-1} + F_1(X_{t-1}) + F_2(X_{t-2}) + DW_t + u_t,$$

where W_t includes a linear time trend and oil price OIL_t.[25] The data has monthly frequency and spans the period between May 1993 and July 2015.[26] Note here that, following the identification strategy of Gonçalves et al. (2021), nonlinear functions F_1 and F_2 are to be understood as not effecting X_t, which is the structural variable. The linear VAR specification of Istrefi and Mouabbi (2018) is recovered by simply assuming $F_1 = F_2 = 0$ prior to estimation. Since they use recursive identification and order the uncertainty measure first, this model too is block-recursive.

I consider a positive shock with intensity $\delta = \sigma_{\epsilon,1}$, where $\sigma_{\epsilon,1}$ is the standard deviation of structural innovations. In this empirical exercise, the relaxation function is given by

$$\rho(z) = \mathbb{I}\left\{|z| \leqslant \frac{1}{4}\right\} \exp\left(1 + \left[|4x|^8 - 1\right]^{-1}\right)$$

and I set $\{0.1, 0.3\}$ to be the cubic spline knots. As 3M3M is a non-negative measure of uncertainty, some care must be taken to make sure that the shocked paths for X_t do not reach negative values. Figure 14 in Appendix C shows that the relaxation function is compatible, and also that the shocked nonlinear paths of X_t with impulse δ and δ' all do not cross below zero.

Figure 7 presents both the linear and nonlinear structural impulse responses obtained. Importantly, even though Istrefi and Mouabbi (2018) estimate a Bayesian VAR model and here

[24] Istrefi and Mouabbi (2018) also provide comparisons with results obtained with the other uncertainty measures, which they comment are all very similar to the ones obtained with 3M3M. Their paper additionally evaluates a number of other highly developed countries.

[25] Inclusion of linear exogenous variables in the semi-nonparametric theoretical framework detail in Section 3 is straightforward as long as one can assume that they are stationary and weakly dependent. The choice of using $p = 2$ is identical to that of the original authors, based on BIC.

[26] I reuse the original data employed by the authors, who kindly shared it upon request, but rescale retail sales (RT_t) so that the level on January 2000 equals 100.

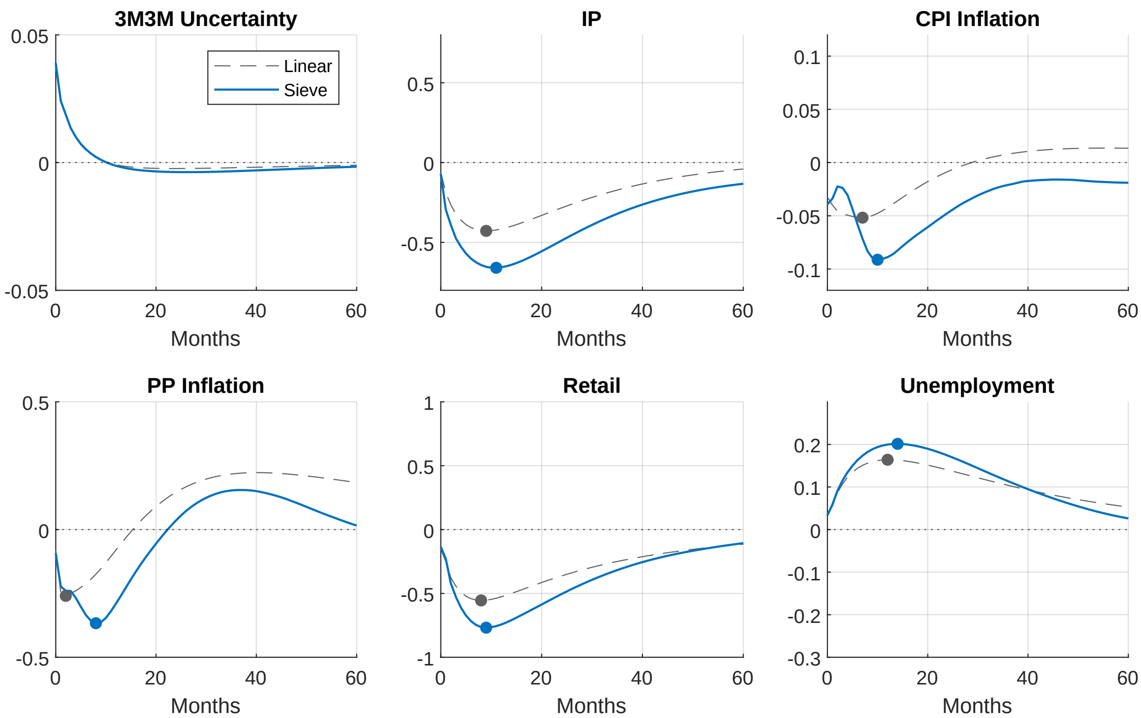


Figure 7: Effect of an unexpected, one-standard-deviation uncertainty shock to US macroeconomic variables. Linear (gray, dashed) and sieve (blue, solid) structural impulse responses. The extreme points of the responses are marked with a dot.

I consider a frequentist vector autoregressive benchmark, the shape of the IRFs is retained, c.f. the median response in the top row of their Figure 4. When uncertainty increases, industrial production drops, and the size and extent of this decrease is intensified in the nonlinear responses. In fact, the sieve IP response reaches a value that is 54% lower than that of the respective linear IRF.[27] A similar behavior holds true for retail sales (38% lower) and unemployment (23% higher), proving that this shock is more profoundly contractionary than suggested by the linear VAR model. Further, CPI and PP inflation both show short-term fluctuations which strengthen the short- and medium-term impact of the shock. CPI and PP nonlinear inflation responses are 76% and 41% stronger than their linear counterpart, respectively. These differences suggest that linear IRFs might be both under-estimating the short-term intensity and misrepresenting long-term persistence of inflation reactions. From another perspective, Nowzohour and Stracca (2020) presented evidence that consumer consumption growth, credit growth and unemployment do not co-move with the policy uncertainty index (EPU) of Baker et al. (2016), but are negatively correlated with financial volatility. Given the strength of nonlinear IRFs, this discrepancy may also suggest that the 3M3M uncertainty measure partially captures the financial channel, too.

The introduction of nonlinear terms in the structural VAR of Istrefi and Mouabbi (2018) thus provides evidence that fundamental impulse response features might otherwise be missed. Indeed, Figure 13 in Appendix C - which plots regression functions of endogenous variables

[27] Figure 15 in Appendix C confirms that this difference is consistent over a range of shock sizes, too.

with respect to X_t - proves that high and low uncertainty levels may have significantly different effects on endogenous economic variables. In particular, at the second lag, tail effects appear to be milder, while at low levels changes in uncertainty have more pronounced impact.

7 Conclusion

This paper studies the application of semi-nonparametric series estimation to the problem of structural impulse response analysis for time series. After first discussing the partial identification model setup, I have used the conditions of system contractivity and stability to derive physical measures of the dependence for nonlinear systems. In turn, these allow to derive primitive conditions under which series estimation can be employed and structural IRFs are consistently estimated. The simulation results prove that this approach is valid in moderate samples and has the added benefit of being robust to misspecification of the nonlinear model components. Finally, two empirical applications showcase the utility in departing from both linear and parametric nonlinear specifications when estimating structural responses.

There are many possible avenues for extending the results I have presented here. A key aspect that I have not touched upon is inference in the form of confidence intervals: the theory of Chen and Christensen (2015) does not encompass uniform inference, and, as such, additional results have to be developed. Indeed, (uniform) confidence bands are necessary to properly quantify the uncertainty of IRF estimates. Belloni et al. (2015) give a uniform asymptotic inference theory, but their derivations are limited to non-dependent data. Li and Liao (2020) and Cattaneo et al. (2022) provide theoretical coupling results that could be exploited in order to handle time series data. Chen and Christensen (2018) give a theory of uniform inference for panel IV setups, which could possibly be generalized to handle nonlinear IRFs. In the spirit of Kang (2021), it would be also important to derive inference results that are uniform in the selection of series terms, as, in practice, a data-driven procedure for selecting K should be used. Studying other sieve spaces, such as neural networks or shape-preserving sieves (Chen, 2007), would also be highly desirable. The latter can be especially useful in contexts where economic knowledge suggests that the nonlinear components of the model are e.g. strictly monotonic increasing or convex. Finally, sharpening of convergence rates used in the main proofs is of independent interest.

References

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.

Caggiano, G., Castelnuovo, E., Colombo, V., and Nodari, G. (2015). Estimating Fiscal Multipliers: News From A Non-linear World. *The Economic Journal*, 125(584):746–776.

Caggiano, G., Castelnuovo, E., and Figueres, J. M. (2017). Economic policy uncertainty and unemployment in the United States: A nonlinear approach. *Economics Letters*, 151:31–34.

Caggiano, G., Castelnuovo, E., and Pellegrino, G. (2021). Uncertainty shocks and the great recession: Nonlinearities matter. *Economics Letters*, 198:109669.

Cattaneo, M. D., Farrell, M. H., and Feng, Y. (2020). Large sample properties of partitioning-based series estimators. *The Annals of Statistics*, 48(3):1718–1741.

Cattaneo, M. D., Masini, R. P., and Underwood, W. G. (2022). Yurinskii's Coupling for Martingales. *Working Paper*.

Chen, X. (2007). Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6, pages 5549–5632. Elsevier.

Chen, X. (2013). Penalized Sieve Estimation and Inference of Seminonparametric Dynamic Models: A Selective Review. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Advances in Economics and Econometrics*, pages 485–544. Cambridge University Press, 1 edition.

Chen, X. and Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188(2):447–465.

Chen, X. and Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics*, 9(1):39–84.

Chen, X., Shao, Q.-M., Wu, W. B., and Xu, L. (2016). Self-normalized Cramér-type moderate deviations under dependence. *The Annals of Statistics*, 44(4):1593–1617.

Chen, X. and Shen, X. (1998). Sieve Extremum Estimates for Weakly Dependent Data. *Econometrica*, 66(2):289.

Chetverikov, D., Santos, A., and Shaikh, A. M. (2018). The Econometrics of Shape Restrictions. *Annual Review of Economics*, 10(1):31–63.

Debortoli, D., Forni, M., Gambetti, L., and Sala, L. (2020). Asymmetric Effects of Monetary Policy Easing and Tightening. *Working Paper.*

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*, volume 20. Springer.

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep Neural Networks for Estimation and Inference. *Econometrica*, 89(1):181–213.

Feng, B. Q. (2003). Equivalence constants for certain matrix norms. *Linear Algebra and its Applications*, 374:247–253.

Forni, M., Gambetti, L., Maffei-Faccioli, N., and Sala, L. (2023a). Nonlinear transmission of financial shocks: Some new evidence. *Journal of Money, Credit and Banking.*

Forni, M., Gambetti, L., and Sala, L. (2023b). Asymmetric effects of news through uncertainty. *Macroeconomic Dynamics*, pages 1–25.

Freyberger, J. and Reeves, B. (2018). Inference under Shape Restrictions. *Working Paper.*

Fuleky, P., editor (2020). *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, volume 52 of *Advanced Studies in Theoretical and Applied Econometrics.* Springer International Publishing, Cham.

Gambetti, L., Maffei-Faccioli, N., and Zoi, S. (2022). Bad News, Good News: Coverage and Response Asymmetries. *Working Paper.*

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Chapman and Hall/CRC.

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

Gourieroux, C. and Jasiak, J. (2005). Nonlinear Innovations and Impulse Responses with Application to VaR Sensitivity. *Annales d'Économie et de Statistique*, pages 1–31.

Gourieroux, C. and Lee, Q. (2023). Nonlinear impulse response functions and local projections. *Working Paper.*

Hamilton, J. D. (1994a). State-space models. *Handbook of Econometrics*, 4:3039–3080.

Hamilton, J. D. (1994b). *Time Series Analysis.* Princeton University Press.

Härdle, W., Lütkepohl, H., and Chen, R. (1997). A Review of Nonparametric Time Series Analysis. *International Statistical Review*, 65(1):49–72.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press, second edition.

Horowitz, J. L. and Lee, S. (2017). Nonparametric estimation and inference under shape restrictions. *Journal of Econometrics*, 201(1):108–126.

Huang, Y., Chen, X., and Wu, W. B. (2014). Recursive Nonparametric Estimation for Time Series. *IEEE Transactions on Information Theory*, 60(2):1301–1312.

Istrefi, K. and Mouabbi, S. (2018). Subjective interest rate uncertainty and the macroeconomy: A cross-country analysis. *Journal of International Money and Finance*, 88:296–313.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.

Kanazawa, N. (2020). Radial basis functions neural networks for nonlinear time series analysis and time-varying effects of supply shocks. *Journal of Macroeconomics*, 64:103210.

Kang, B. (2021). Inference In Nonparametric Series Estimation with Specification Searches for the Number of Series Terms. *Econometric Theory*, 37(2):311–345.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press, Cambridge.

Kilian, L. and Vega, C. (2011). Do energy prices respond to us macroeconomic news? a test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics*, 93(2):660–671.

Kilian, L. and 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., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.

Lanne, M. and Nyberg, H. (2023). Nonparametric Impulse Response Analysis in Changing Macroeconomic Conditions. *Working Paper*.

Li, J. and Liao, Z. (2020). Uniform nonparametric inference for time series. *Journal of Econometrics*, page 14.

Li, Q. and Racine, J. S. (2009). *Nonparametric econometric methods*. Emerald Group Publishing.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York : Springer, Berlin.

Movahedifar, M. and Dickhaus, T. (2023). On the closed-loop Volterra method for analyzing time series. *Working Paper*.

Nowzohour, L. and Stracca, L. (2020). More than a feeling: Confidence, uncertainty, and macroeconomic fluctuations. *Journal of Economic Surveys*, 34(4):691–726.

Pellegrino, G. (2021). Uncertainty and monetary policy in the US: A journey into nonlinear territory. *Economic Inquiry*, 59(3):1106–1128.

Pötscher, B. M. and Prucha, I. (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Science & Business Media.

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

Ramey, V. A. and 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. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, 94(4):1055–1084.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48.

Sirotko-Sibirskaya, N., Franz, M. O., and Dickhaus, T. (2020). Volterra bootstrap: Resampling higher-order statistics for strictly stationary univariate time series. *Working Paper*.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.

Tenreyro, S. and Thwaites, G. (2016). Pushing on a string: US monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics*, 8(4):43–74.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press.

Tong, H. (1990). *Non-linear Time Series: a Dynamical System Approach*. Oxford University Press.

Tropp, J. A. (2012). User-Friendly Tail Bounds for Sums of Random Matrices. *Foundations of Computational Mathematics*, 12(4):389–434.

Tsay, R. S. and Chen, R. (2018). *Nonlinear Time Series Analysis*, volume 891. John Wiley & Sons.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York ; London.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2):207–226.

Wu, W. B., Huang, Y., and Huang, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications*, 120(12):2412–2431.

Appendix

A Preliminaries

Matrix Norms. Let

$$\|A\|_r := \max\left\{\|Ax\|_r \,\middle|\, \|x\|_r \leqslant 1\right\}$$

be the r-operator norm of matrix $A \in \mathbb{C}^{d_1\times d_2}$. The following Theorem establishes the equivalence between different operator norms as well as the compatibility constants.

Theorem A.1 (Feng (2003))**.** *Let $1 \leqslant p, q \leqslant \infty$. Then for all $A \in \mathbb{C}^{d_1\times d_2}$,*

$$\|A\|_p \leqslant \lambda_{p,q}(d_1)\lambda_{q,p}(d_2)\|A\|_q,$$

where

$$\lambda_{a,b}(d) := \begin{cases} 1 & \text{if } a \geqslant b, \\ d^{1/a-1/b} & \text{if } a < b. \end{cases}$$

This norm inequality is sharp.

In particular, if $p > q$ then it holds

$$\frac{1}{(d_2)^{1/q-1/p}}\|A\|_p \leqslant \|A\|_q \leqslant (d_1)^{1/q-1/p}\|A\|_p.$$

B Proofs

B.1 GMC Conditions and Proposition 3.1

Lemma B.1. *Assume that $\{\epsilon_t\}_{t\in\mathbb{Z}}$, $\epsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_Z}$ are i.i.d., and $\{Z_t\}_{t\in\mathbb{Z}}$ is generated according to*

$$Z_t = G(Z_{t-1}, \epsilon_t),$$

where $Z_t \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$ and G is a measurable function. If either

(a) Contractivity conditions (11)-(12) hold, $\sup_{t\in\mathbb{Z}}\|\epsilon_t\|_{L^r} < \infty$ and $\|G(\overline{z}, \overline{\epsilon})\| < \infty$ for some $(\overline{z}, \overline{\epsilon}) \in \mathcal{Z} \times \mathcal{E}$;

(b) Stability conditions (13)-(14) hold, $\sup_{t\in\mathbb{Z}}\|\epsilon_t\|_{L^r} < \infty$ and $\|\partial G/\partial Z\| \leqslant M_Z < \infty$;

then

$$\sup_t \|Z_t\|_{L^r} < \infty \quad w.p.1.$$

Proof.

(a) In a first step, we show that, given event $\omega \in \Omega$, realization $Z_t(\omega)$ is unique with probability one. To do this, introduce initial condition $z_\circ$ for $\ell > 1$ such that $z_\circ \in \mathcal{Z}$ and $\|z_\circ\| < \infty$. Define

$$Z_t^{(-\ell)}(\omega) = G^{(\ell)}(y_\circ, \epsilon_{t-\ell+1:t}(\omega)).$$

Further, let $Z_t^{\prime(-\ell)}$ be the realization with initial condition $z'_\circ \neq z_\circ$ and innovation realizations $\epsilon_{t-\ell+1:t}(\omega)$. Note that

$$\left\| Z_t^{(-\ell)}(\omega) - Z_t^{\prime(-\ell)}(\omega) \right\| \leqslant C_Z^\ell \left\| z_\circ - z'_\circ \right\|,$$

which goes to zero as $\ell \to \infty$. Therefore, if we set $Z_t(\omega) := \lim_{\ell\to\infty} Z_t^{(-\ell)}(\omega)$, $Z_t(\omega)$ is unique with respect to the choice of $z_\circ$ w.p.1. A similar recursion shows that

$$\left\| Z_t^{(-\ell)}(\omega) \right\| \leqslant C_Z^\ell \left\| z_\circ \right\| + \sum_{k=0}^{\ell-1} C_Z^k C_\epsilon \left\| \epsilon_{t-k}(\omega) \right\|.$$

By norm equivalence, this implies

$$\begin{aligned} \left\| Z_t^{(-\ell)} \right\|_{L^r} &\leqslant C_Z^\ell \left\| z_\circ \right\|_r + \sum_{k=0}^{\ell-1} C_Z^k C_\epsilon \left\| \epsilon_{t-k} \right\|_{L^r} \\ &\leqslant C_Z^\ell \left\| z_\circ \right\|_r + (1 - C_Z)^{-1} C_\epsilon \sup_{t\in\mathbb{Z}} \|\epsilon_t\|_{L^r} < \infty, \end{aligned}$$

and taking the limit $\ell \to \infty$ proves the claim.

(b) Consider again distinct initial conditions $z'_\circ \neq z_\circ$ and innovation realizations $\epsilon_{t-\ell+1:t}(\omega)$, yielding $Z_t^{\prime(-\ell)}(\omega)$ and $Z_t^{(-\ell)}(\omega)$, respectively. We may use the contraction bound derived in the proof of Proposition 3.1 (b) below, that is,

$$\left\| Z_t^{(-\ell)}(\omega) - Z_t^{\prime(-\ell)}(\omega) \right\|_r \leqslant C_Z^\ell C_2 \|z_\circ - z'_\circ\|_r,$$

where $C_2 > 0$ is a constant. With trivial adjustments, the uniqueness and limit arguments used for (a) above apply here too.

□

Proof of Proposition 3.1.

(a) By assumption it holds that for all $(z, z') \in \mathcal{Z} \times \mathcal{Z}$ and $(e, e') \in \mathcal{E} \times \mathcal{E}$

$$\|G(z, \epsilon) - G(z', \epsilon')\| \leqslant C_Z \|z - z'\| + C_\epsilon \|e - e'\|$$

holds, where $0 \leqslant C_Z < 1$ and $0 \leqslant C_\epsilon < \infty$. The equivalence of norms directly generalizes this inequality to any r-norm for $r > 2$. We study $\|Z_{t+h} - Z'_{t+h}\|_r$ where Z'_{t+h} is constructed with a time-t perturbation of the history of Z_{t+h}. Therefore, for any given t and $h \leqslant 1$ it holds that

$$\begin{aligned} \left\| Z_{t+h} - G^{(h)}(Z'_t, \epsilon_{t+1:t+h}) \right\|_r &\leqslant C_Z \|G^{(h-1)}(Z_t, \epsilon_{t+1:t+h-1}) - G^{(h-1)}(Z'_t, \epsilon_{t+1:t+h-1})\|_r \\ &\leqslant C_Z^h \|Z_t - Z'_t\|_r, \end{aligned}$$

since sequence $\epsilon_{t+1:t+h}$ is common between Z_{t+h} and Z'_{t+h}. Clearly then

$$\left\| Z_{t+h} - G^{(h)}(Z'_t, \epsilon_{t+1:t+h}) \right\|_r \leqslant 2\|Z_t\|_r \exp(-\gamma h)$$

for $\gamma = -\log(C_Z)$. Letting $a = 2\|Z_t\|_r$ and shifting time index t backward by h, since $\sup_t\|Z_t\|_{L^r} < \infty$ w.p.1 from Lemma B.1 the result for L^r follows with $\tau = 1$.

(b) Proceed similar to (a), but notice that now we must handle cases of steps $1 \leqslant h < h^*$. Consider iterate $h^* + 1$, for which

$$
\begin{aligned}
\left\| Z_{t+h+1} - G^{(h+1)}(Z'_t, \epsilon_{t+1:t+h+1}) \right\|_r &\leqslant C_Z\|G^{(h)}(G(Z_t, \epsilon_{t+1}), \epsilon_{t+2:t+h}) - G^{(h)}(G(Z'_t, \epsilon_{t+1}), \epsilon_{t+2:t+h})\|_r \\
&\leqslant C_Z^h\|G(Z_t, \epsilon_{t+1}) - G(Z'_t, \epsilon_{t+1})\|_r \\
&\leqslant C_Z^h M_Z\|Z_t - Z'_t\|_r
\end{aligned}
$$

by the mean value theorem. Here we may assume that $M_Z \geqslant 1$ otherwise we would fall under case (a), so that $M_Z \leqslant M_Z^2 \leqslant \ldots \leqslant M_Z^{h^*-1}$. More generally,

$$
\left\| Z_{t+h+1} - G^{(h+1)}(Z'_t, \epsilon_{t+1:t+h+1}) \right\|_r \leqslant C_Z^{j(h)} \max\{M_Z^{h^*-1}, 1\}\|Z_t - Z'_t\|_r
$$

for $j(h) := \lfloor h/h^* \rfloor$. Result (b) then follows by noting that $j(h) \geqslant h/h^* - 1$ and then proceeding as in (a) to derive GMC coefficients.

□

Companion and Lagged Vectors. The assumption of GMC for a process translates naturally to vectors that are composed of stacked lags of realizations. This, for example, is important in the discussion of Section 3 when imposing Assumption 9, since one needs that series regressors $\{W_{2t}\}_{t\in\mathbb{Z}}$ be GMC.

Recall that $W_{2t} = (X_t, X_{t-1}, \ldots, X_{t-p}, Y_{t-1}, \ldots, Y_{t-p}, \epsilon_{1t})$. Here we shall reorder this vector slightly to be

$$W_{2t} = (X_t, X_{t-1}, Y_{t-1}, \ldots, X_{t-p}, Y_{t-p}, \epsilon_{1t}).$$

For $h > 0$ and $1 \leqslant l \leqslant h$, let $Z'_{t+j} := \Phi^{(l)}(Z'_t, \ldots, Z'_{t-p}; \epsilon_{t+1:t+j})$ be the a perturbed version of Z_t, where $Z'_t, \ldots, Z'_{t-p}$ are taken from an independent copy of $\{Z_t\}_{t\in\mathbb{Z}}$. Define

$$W'_{2t} = (X'_t, X'_{t-1}, Y'_{t-1}, \ldots, X'_{t-p}, Y'_{t-p}, \epsilon_{1t}).$$

Using Minkowski's inequality

$$
\begin{aligned}
\|W_{2t+h} - W'_{2t+h}\|_{L^r} &\leqslant \|X_{t+h} - X'_{t+h}\|_{L^r} + \sum_{j=1}^{p}\|Z_{t+h-j} - Z'_{t+h-j}\|_{L^r} \\
&\leqslant \sum_{j=0}^{p}\|Z_{t+h-j} - Z'_{t+h-j}\|_{L^r},
\end{aligned}
$$

thus, since $p > 0$ is fixed finite,

$$\sup_t\|W_{2t+h} - W'_{2t+h}\|_{L^r} \leqslant \sum_{j=0}^{p} \Delta_r(h-j) \leqslant (p+1)\, a_{1Z} \exp(-a_{2Z} h).$$

Above, a_{1Z} and a_{2Z} are the GMC coefficients of $\{Z_t\}_{t\in\mathbb{Z}}$.

B.2 Lemma 3.1 and Matrix Inequalities under Dependence

In order to prove Lemma 3.1, the idea is to modify the approach of Chen and Christensen (2015), which relies on Berbee's Lemma and an interlaced coupling, to handle variables with physical dependence. Chen et al. (2016) provide an example on how to achieve this when working with self-normalized sums. In what follows I modify their ideas to work with random dependent matrices.

First of all, I recall below a Bernstein-type inequality for independent random matrices of Tropp (2012).

Theorem B.1. *Let $\{\Xi_i\}_{i=1}^n$ be a finite sequence of independent random matrices with dimensions $d_1 \times d_2$. Assume $\mathbb{E}[\Xi_i] = 0$ for each i and $\max_{1\leqslant i\leqslant n}\|\Xi_i\| \leqslant R_n$ and define*

$$\varsigma_n^2 := \max\left\{\left\|\sum_{i=1}^n \mathbb{E}\left[\Xi_{i,n}\Xi_{j,n}'\right]\right\|, \left\|\sum_{i=1}^n \mathbb{E}\left[\Xi_{i,n}'\Xi_{j,n}\right]\right\|\right\}.$$

Then for all $z \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_i\right\| \geqslant z\right) \leqslant (d_1+d_2)\exp\left(\frac{-z^2/2}{nq\varsigma_n^2 + qR_n z/3}\right).$$

The main exponential matrix inequality due to Chen and Christensen (2015), Theorem 4.2 is as follows.

Theorem B.2. *Let $\{X_i\}_{i\in\mathbb{Z}}$ where $X_i \in \mathcal{X}$ be a β-mixing sequence and let $\Xi_{i,n} = \Xi_n(X_i)$ for each i where $\Xi_n : \mathcal{X} \mapsto \mathbb{R}^{d_1\times d_2}$ be a sequence of measurable $d_1 \times d_2$ matrix-valued functions. Assume that $\mathbb{E}[\Xi_{i,n}] = 0$ and $\|\Xi_{i,n}\| \leqslant R_n$ for each i and define*

$$S_n^2 := \max\left\{\mathbb{E}\left[\|\Xi_{i,n}\Xi_{j,n}'\|\right], \mathbb{E}\left[\|\Xi_{i,n}'\Xi_{j,n}\|\right]\right\}.$$

Let $1 \leqslant q \leqslant n/2$ be an integer and let $I_\bullet = q\lfloor n/q\rfloor, \ldots, n$ when $q\lfloor n/q\rfloor < n$ and $I_\bullet = \varnothing$ otherwise. Then, for all $z \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_{i,n}\right\| \geqslant 6z\right) \leqslant \frac{n}{q}\beta(q) + \mathbb{P}\left(\left\|\sum_{i\in I_\bullet} \Xi_{i,n}\right\| \geqslant z\right) + 2(d_1+d_2)\exp\left(\frac{-z^2/2}{nqS_n^2 + qR_n z/3}\right),$$

where $\|\sum_{i\in I_\bullet}\Xi_{i,n}\| := 0$ whenever $I_\bullet = \varnothing$.

To fully extend Theorem B.2 to physical dependence, I will proceed in steps. First, I derive a similar matrix inequality by directly assuming that random matrices $\Xi_{i,n}$ have physical dependence coefficient $\Delta_r^\Xi(h)$. In the derivations I will use that

$$\frac{1}{(d_2)^{1/2-1/r}}\|A\|_r \leqslant \|A\|_2 \leqslant (d_1)^{1/2-1/r}\|A\|_r.$$

for $r \geqslant 2$.

Theorem B.3. *Let $\{\epsilon_j\}_{j\in\mathbb{Z}}$ be a sequence of i.i.d. variables and let $\{\Xi_{i,n}\}_{i=1}^n$,*

$$\Xi_{i,n} = G_n^{\Xi}(\ldots,\epsilon_{i-1},\epsilon_i)$$

for each i, where $\Xi_n : \mathcal{X} \mapsto \mathbb{R}^{d_1\times d_2}$, be a sequence of measurable $d_1 \times d_2$ matrix-valued functions. Assume that $\mathbb{E}[\Xi_{i,n}] = 0$ and $\|\Xi_{i,n}\| \leqslant R_n$ for each i and define

$$S_n^2 := \max\left\{\mathbb{E}\left[\|\Xi_{i,n}\Xi_{j,n}'\|\right], \mathbb{E}\left[\|\Xi_{i,n}'\Xi_{j,n}\|\right]\right\}.$$

Additionally assume that $\|\Xi_{i,n}\|_{L^r} < \infty$ for $r > 2$ and define the matrix physical dependence measure $\Delta_r^{\Xi}(h)$ as

$$\Delta_r^{\Xi}(h) := \max_{1\leqslant i\leqslant n}\left\|\,\Xi_{i,n} - \Xi_{i,n}^{h*}\,\right\|_{L^r},$$

where $\Xi_{i,n}^{h*} := G_n^{\Xi}(\ldots,\epsilon_{i-h-1}^*,\epsilon_{i-h}^*,\epsilon_{i-h+1},\ldots,\epsilon_{i-1},\epsilon_i)$ for independent copy $\{\epsilon_j^*\}_{j\in\mathbb{Z}}$. Let $1 \leqslant q \leqslant n/2$ be an integer and let $I_\bullet = q\lfloor n/q\rfloor,\ldots,n$ when $q\lfloor n/q\rfloor < n$ and $I_\bullet = \varnothing$ otherwise. Then, for all $z \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_{i,n}\right\| \geqslant 6z\right) \leqslant \frac{n^{r+1}}{q^r(d_2)^{r/2-1}z^r}\,\Delta_r^{\Xi}(q) + \mathbb{P}\left(\left\|\sum_{i\in I_\bullet}\Xi_{i,n}\right\| \geqslant z\right) + 2(d_1+d_2)\exp\left(\frac{-z^2/2}{nqS_n^2 + qR_nz/3}\right),$$

where $\|\sum_{i\in I_\bullet}\Xi_{i,n}\| := 0$ whenever $I_\bullet = \varnothing$.

Proof. To control dependence, we can adapt the interlacing block approach outlined by Chen et al. (2016). To interlace the sum, split it into

$$\sum_{i=1}^n \Xi_{i,n} = \sum_{j\in K_e} J_k + \sum_{j\in J_o} W_k + \sum_{i\in I_\bullet}\Xi_{i,n},$$

where $W_j := \sum_{i=q(j-1)+1}^{qj}\Xi_{i,n}$ for $j = 1,\ldots,\lfloor n/q\rfloor$ are the blocks, $I_\bullet := \{q\lfloor n/q\rfloor + 1,\ldots,n\}$ if $q\lfloor n/q\rfloor < n$ and J_e and J_o are the subsets of even and odd numbers of $\{1,\ldots,\lfloor n/q\rfloor\}$, respectively. For simplicity define $J = J_e \cup J_o$ as the set of block indices and let

$$W_j^\dagger := \mathbb{E}\big[\,W_j \mid \epsilon_\ell,\, q(j-2)+1 \leqslant \ell \leqslant qj\,\big].$$

Note that by construction $\{W_j^\dagger\}_{j\in J_e}$ are independent and also $\{W_j^\dagger\}_{j\in J_o}$ are independent. Using the triangle inequality we find

$$\begin{aligned}
\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_{i,n}\right\| \geqslant 6z\right) &\leqslant \mathbb{P}\left(\left\|\sum_{j\in J}(W_j - W_j^\dagger)\right\| + \left\|\sum_{j\in J}W_j^\dagger\right\| + \left\|\sum_{i\in I_\bullet}\Xi_{i,n}\right\| \geqslant 6z\right)\\
&\leqslant \mathbb{P}\left(\left\|\sum_{j\in J}(W_j - W_j^\dagger)\right\| \geqslant z\right) + \mathbb{P}\left(\left\|\sum_{j\in J_e}W_j^\dagger\right\| \geqslant z\right)\\
&\quad + \mathbb{P}\left(\left\|\sum_{j\in J_o}W_j^\dagger\right\| \geqslant z\right) + \mathbb{P}\left(\left\|\sum_{i\in I_\bullet}\Xi_{i,n}\right\| \geqslant z\right)\\
&= I + II + III + IV.
\end{aligned}$$

We keep term IV as is. As in the proof of Chen and Christensen (2015), terms II and III consist of sums of independent matrices, where each $W_j^\dagger$ satisfies $\|W_j^\dagger\| \leqslant qR_n$ and

$$\max\left\{\mathbb{E}\left[\|W_j^\dagger W_j^{\dagger\prime}\|\right], \mathbb{E}\left[\|W_j^{\dagger\prime} W_j^\dagger\|\right]\right\} \leqslant qS_n^2.$$

Then, using the exponential matrix inequality of Tropp (2012),

$$\mathbb{P}\left(\left\|\sum_{j\in J_e} W_k^\dagger\right\| \geqslant z\right) \leqslant (d_1+d_2)\exp\left(\frac{-z^2/2}{nqS_n^2 + qR_n z/3}\right).$$

The same holds for the sum over J_o. Finally, we use the physical dependence measure Δ_r^Ξ to bound I. Start with the union bound to find

$$\begin{aligned}\mathbb{P}\left(\left\|\sum_{j\in J}(W_j - W_j^\dagger)\right\| \geqslant z\right) &\leqslant \mathbb{P}\left(\sum_{j\in J}\left\|W_j - W_j^\dagger\right\| \geqslant z\right)\\ &\leqslant \frac{n}{q}\,\mathbb{P}\left(\left\|W_j - W_j^\dagger\right\| \geqslant \frac{q}{n}\,z\right),\end{aligned}$$

where we have used that $\lfloor n/q\rfloor \leqslant n/q$. Since W_j and $W_j^\dagger$ differ only over a σ-algebra that is q steps in the past, by assumption

$$\left\|W_j - W_j^\dagger\right\|_{L^r} \leqslant q\,\Delta_r^\Xi(q),$$

which implies, by means of the rth moment inequality,

$$\mathbb{P}\left(\left\|W_j - W_j^\dagger\right\| \geqslant \frac{q}{n}\,z\right) \leqslant \mathbb{P}\left((d_2)^{1/r-1/2}\left\|W_j - W_j^\dagger\right\|_r \geqslant \frac{q}{n}\,z\right) \leqslant \frac{n^r}{q^{r-1}(d_2)^{r/2-1}z^r}\,\Delta_r^\Xi(q).$$

where $(d_2)^{1/r-1/2}$ is the operator norm equivalence constant such that $\|\cdot\| \geqslant (d_2)^{1/r-1/2}\|\cdot\|_r$ (Feng, 2003). Therefore,

$$\mathbb{P}\left(\left\|\sum_{j\in J}(W_j - W_j^\dagger)\right\| \geqslant z\right) \leqslant \frac{n^{r+1}}{q^r(d_2)^{r/2-1}z^r}\,\Delta_r^\Xi(q)$$

as claimed. □

Notice that the first term in the bound is weaker than that derived by Chen and Christensen (2015). The β-mixing assumption and Berbee's Lemma give strong control over the probability $\mathbb{P}(\|\sum_{j\in J}(W_j - W_j^\dagger)\| \geqslant z)$. In contrast, assuming physical dependence means we have to explicitly handle a moment condition. One might think of sharpening Theorem B.3 by sidestepping the rth moment inequality (c.f. avoiding Chebyshev's inequality in concentration results), but I do not explore this approach here.

The second step is to map the physical dependence of a generic vector time series $\{X_i\}_{i\in\mathbb{Z}}$ to matrix functions.

Proposition B.1. *Let $\{X_i\}_{i\in\mathbb{Z}}$ where $X_i = G(\ldots,\epsilon_{i-1},\epsilon_i) \in \mathcal{X}$ for $\{\epsilon_j\}_{j\in\mathbb{Z}}$ i.i.d. be a sequence* *with finite rth moment, where $r > 0$, and functional physical dependence coefficients*

$$\Delta_r(h) = \sup_i \left\| X_{i+h} - G^{(h)}(X_i^*, \epsilon_{i+1:i+h}) \right\|_{L^r}$$

for $h \geqslant 1$. Let $\Xi_{i,n} = \Xi_n(X_i)$ for each i where $\Xi_n : \mathcal{X} \mapsto \mathbb{R}^{d_1 \times d_2}$ be a sequence of measurable $d_1 \times d_2$ matrix-valued functions such that $\Xi_n = (v_1, \ldots, v_{d_2})$ for $v_\ell \in \mathbb{R}^{d_1}$. If $\|\Xi_{i,n}\|_{L^r} < \infty$ and

$$C_{\Xi,\ell} := \sup_{x \in \mathcal{X}} \|\nabla v_\ell(x)\| \leqslant C_\Xi < \infty,$$

then matrices $\Xi_{i,n}$ have physical dependence coefficients

$$\Delta_r^\Xi(h) = \sup_i \left\| \Xi_{i,n} - \Xi_{i,n}^{h*} \right\|_{L^r} \leqslant \sqrt{d_1} \left(\frac{d_2}{d_1} \right)^{1/r} C_\Xi \, \Delta_r(h),$$

where $\Xi_{i,n}^{h*} = \Xi_n(G^{(h)}(X_i', \epsilon_{i+1:i+h}))$.

Proof. To derive the bound, we use $\Xi_n(X_i)$ and $\Xi_n(X_i^{h*})$ in place of $\Xi_{i,n}$ and $\Xi_{i,n}^{h*}$, respectively, where $X_i^{h*} = G^{(h)}(X_i^*, \epsilon_{i+1:i+h})$. First we move from studying the operator r-norm (recall, $r > 2$) to the Frobenius norm,

$$\left\| \Xi_n(X_i) - \Xi_n(X_i^{h*}) \right\|_r \leqslant (d_2)^{1/2 - 1/r} \left\| \Xi_n(X_i) - \Xi_n(X_i^{h*}) \right\|_F.$$

where as intermediate step we use the 2-norm. Let $\Xi_n = (v_1, \ldots, v_{d_2})$ for $v_\ell \in \mathbb{R}^{d_1}$ and $\ell \in 1, \ldots, d_2$, so that

$$\|\Xi_n\|_F = \sqrt{\sum_{\ell=1}^{d_2} \|v_\ell\|^2}$$

where $v_\ell = (v_{\ell 1}, \ldots, v_{\ell d_1})'$. Since $v_\ell : \mathcal{X} \mapsto \mathbb{R}^{d_1}$ are vector functions, the mean value theorem gives that

$$\left\| \Xi_n(X_i) - \Xi_n(X_i^{h*}) \right\|_F \leqslant \sqrt{\sum_{\ell=1}^{d_2} C_{\Xi,\ell}^2 \, \|X_i - X_i^{h*}\|^2} \leqslant \sqrt{d_2} \, C_\Xi \, \|X_i - X_i^{h*}\|.$$

Combining results and moving from the vector r-norm to the 2-norm yields

$$\left\| \Xi_n(X_i) - \Xi_n(X_i^{h*}) \right\|_r \leqslant (d_2)^{1-1/r} (d_1)^{1/2 - 1/r} \, C_\Xi \, \|X_i - X_i^{h*}\|_r.$$

The claim involving the L^r norm follows immediately. $\square$

The following Corollary, which specifically handles matrix functions defined as outer products of vector functions, is immediate and covers the setups of series estimation.

Corollary B.1. *Under the conditions of Proposition B.1, if*

$$\Xi_n(X_i) = \xi_n(X_i) \xi_n(X_i)' + Q_n$$

where $\xi_n : \mathcal{X} \mapsto \mathbb{R}^d$ *is a vector function and* $Q_n \in \mathbb{R}^{d\times d}$ *is nonrandom matrix, then*

$$\Delta_r^{\Xi}(h) \leqslant d^{3/2-2/r}\, C_\xi\, \Delta_r(h),$$

where $C_\xi := \sup_{x\in\mathcal{X}} \|\nabla \xi_n(x)\| < \infty$.

Proof. Matrix Q_n cancels out since it is nonrandom and appears in both $\Xi_n(X_i)$ and $\Xi_n(X_i^{h*})$. Since $\Xi_n(X_i)$ is square, the ratio of row to column dimensions simplifies. □

The following Corollaries to Theorem B.3 can now be derived in a straightforward manner.

Corollary B.2. *Under the conditions of Theorem B.3 and Proposition B.1, for all* $z \geqslant 0$

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_{i,n}\right\| \geqslant 6z\right) \leqslant \frac{n^{r+1}}{q^r z^r}(d_2)^{2-(r/2+1/r)}(d_1)^{1/2-1/r} C_\Xi\, \Delta_r(q) + \mathbb{P}\left(\left\|\sum_{i\in I_\bullet} \Xi_{i,n}\right\| \geqslant z\right) + 2(d_1+d_2)\exp\left(\frac{-z^2/2}{nqS_n^2 + qR_n z/3}\right).$$

where $\Delta_r(\cdot)$ *if the functional physical dependence coefficient of* X_i.

Corollary B.3. *Under the conditions of Theorem B.3 and Proposition B.1, if* $q = q(n)$ *is chosen such that*

$$\frac{n^{r+1}}{q^r}(d_2)^{2-(r/2+1/r)}(d_1)^{1/2-1/r} C_\Xi\, \Delta_r(q) = o(1)$$

and $R_n\sqrt{q\log(d_1+d_2)} = o(S_n\sqrt{n})$ *then*

$$\left\|\sum_{i=1}^n \Xi_{i,n}\right\| = O_P\left(S_n\sqrt{nq\log(d_1+d_2)}\right).$$

This result is almost identical to Corollary 4.2 in Chen and Christensen (2015), with the only adaptation of using Theorem B.3 as a starting point. Condition $R_n\sqrt{q\log(d_1+d_2)} = o(S_n\sqrt{n})$ is simple to verify by assuming, e.g., $q = o(n/\log(n))$ since $\log(d_1+d_2) \lesssim \log(K)$ and $K = o(n)$.

Note that when $d_1 = d_2 \equiv K$, which is the case of interest in the series regression setup, the first condition in Corollary B.3 reduces to

$$K^{5/2-(r/2+2/r)}\, C_\Xi\, \Delta_r(q) = o(1),$$

which also agrees with the rate of Corollary B.1. Assumption 7(i) and a compact domain further allow to explicitly bound factor C_Ξ by

$$C_\Xi \lesssim K^{\omega_2},$$

so that the required rate becomes

$$K^\rho\, \Delta_r(q) = o(1), \quad \text{where} \quad \rho := \frac{3}{2} - \frac{r}{2} + \omega_2.$$

Proof of Lemma 3.1. The proof follows from Corollary B.3 by the same steps of the proof of Lemma 2.2 in Chen and Christensen (2015). Simply take $\Xi_{i,n} = n^{-1}(\widetilde{(b)}^K_\pi(X_i)\widetilde{(b)}^K_\pi(X_i)' - I_K)$ and note that $R_n \leqslant n^{-1}(1 + \zeta^2_{K,n}\lambda^2_{K,n})$ and $S_n \leqslant n^{-2}(1 + \zeta^2_{K,n}\lambda^2_{K,n})$. $\square$

For Lemma 3.1 to hold under GMC assumptions a valid choice for $q(n)$ is

$$q(n) = \gamma^{-1} \log(K^\rho n^{r+1})$$

where γ as in Proposition 3.1. This is due to

$$\begin{aligned} \left(\frac{n}{q}\right)^{r+1} qK^\rho \Delta_r(q) &\lesssim \frac{n^{r+1}}{q^r} K^\rho \exp(-\gamma q) \\ &\lesssim \frac{n^{r+1}K^\rho}{\log(K^\rho n^{r+1})^r} (K^\rho n^{r+1})^{-1} \\ &= \frac{1}{\log(K^\rho n^{r+1})^r} = o(1). \end{aligned}$$

Note then that, if $\lambda_{K,n} \lesssim 1$ and $\zeta_{K,n} \lesssim \sqrt{K}$, since

$$\zeta_{K,n}\lambda_{K,n}\sqrt{\frac{q \log K}{n}} \lesssim \sqrt{\frac{K \log(K^\rho n^{r+1}) \log(K)}{n}} \lesssim \sqrt{\frac{K \log(n^{\rho+r+2}) \log(n)}{n}} \lesssim \sqrt{\frac{K \log(n)^2}{n}},$$

to satisfy Assumption 8 we may assume $\sqrt{K \log(n)^2/n} = o(1)$ as in Remark 2.3 of Chen and Christensen (2015) for the case of exponential β-mixing regressors.

B.3 Theorem 3.2

Before delving into the proof of Theorem 3.2, note that we can decompose $\widehat{\Pi}_2 - \Pi_2$ as

$$\widehat{\Pi}_2 - \Pi_2 = (\widehat{\Pi}_2 - \widehat{\Pi}^*_2) + (\widehat{\Pi}^*_2 - \widetilde{\Pi}_2) + (\widetilde{\Pi}_2 - \Pi_2),$$

where $\widetilde{\Pi}_2$ is the projection of Π_2 onto the linear space spanned by the sieve. The last two terms can be handled directly with the theory developed by Chen and Christensen (2015). Specifically, their Lemma 2.3 controls the second term (variance term), while Lemma 2.4 handles the third term (bias term). This means here we can focus on the first term, which is due to using generated regressors $\widehat{\epsilon}_{1t}$ in the second step.

Since $\widehat{\Pi}_2$ can be decomposed in d_Y rows of semi-nonparametric coefficients, i.e.,

$$Y_t = \begin{bmatrix} \pi_{2,1} \\ \vdots \\ \pi_{2,d_Y} \end{bmatrix} W_{2t} + \widetilde{u}_{2t},$$

we further reduce to the scalar case. Let π_2 be any row of Π_2 and, with a slight abuse of notation, Y the vector of observations of the component of Y_t of the same row, so that one may write

$$\widehat{\pi}_2(x) - \widehat{\pi}^*_2(x) = \widetilde{b}^K_\pi(x)\big(\widehat{\widetilde{B}}{}'_\pi \widehat{\widetilde{B}}_\pi\big)^-\big(\widehat{\widetilde{B}}_\pi - \widetilde{B}_\pi\big)'Y + \widetilde{b}^K_\pi(x)\left[\big(\widehat{\widetilde{B}}{}'_\pi \widehat{\widetilde{B}}_\pi\big)^- - \big(\widetilde{B}'_\pi \widetilde{B}_\pi\big)^-\right]\widetilde{B}'_\pi Y$$

$$= I + II$$

where $\widetilde{b}_\pi^K(x) = \Gamma_{B,2}^{-1/2} b_\pi^K(x)$ is the orthonormalized sieve according to $\Gamma_{B,2} := \mathbb{E}[b_\pi^K(W_{2t}) b_\pi^K(W_{2t})']$, $\widetilde{B}_\pi$ is the *infeasible* orthonormalized design matrix (involving ϵ_{1t}) and $\widehat{\widetilde{B}}_\pi$ is *feasible* orthonormalized design matrix (involving $\widehat{\epsilon}_{1t}$). In particular, note that

$$\widehat{B}_\pi = B_\pi + R_n, \quad \text{where} \quad R_n := \begin{bmatrix} 0 & & 0 & \widehat{\epsilon}_{11} - \epsilon_{11} \\ \vdots & \dots & \vdots & \vdots \\ 0 & & 0 & \widehat{\epsilon}_{1n} - \epsilon_{1n} \end{bmatrix} \in \mathbb{R}^{n \times K},$$

which implies $\widehat{\widetilde{B}}_\pi - \widetilde{B}_\pi = R_n \Gamma_{B,2}^{-1/2} =: \widetilde{R}_n$.

The next Lemma provides a bound for the difference $(\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi / n) - (\widetilde{B}_\pi' \widetilde{B}_\pi / n)$ that will be useful in the proof of Theorem 3.1 below.

Lemma B.2. *Under the setup of Theorem 3.1, it holds*

$$\|(\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi / n) - (\widetilde{B}_\pi' \widetilde{B}_\pi / n)\| = O_P(\sqrt{K/n}).$$

Proof. Using the expansion $\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi = \widetilde{B}_\pi' \widetilde{B}_\pi + (\widetilde{B}_\pi' \widetilde{R}_n + \widetilde{R}_n' \widetilde{B}_\pi) + \widetilde{R}_n' \widetilde{R}_n$, one immediately finds that

$$\|(\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi / n) - (\widetilde{B}_\pi' \widetilde{B}_\pi / n)\| \leqslant 2 \|\widetilde{B}_\pi' \widetilde{R}_n / n\| + \|\widetilde{R}_n' \widetilde{R}_n / n\|.$$

The second right-hand side factor satisfies $\|\widetilde{R}_n' \widetilde{R}_n / n\| \leqslant \lambda_{K,n}^2 \|R_n' R_n / n\|$. Moreover,

$$\begin{aligned} \|R_n' R_n / n\| &= \left\| \frac{1}{n} \sum_{t=1}^n (\widehat{\epsilon}_{1t} - \epsilon_{1t})^2 \right\| \\ &= \left\| \frac{1}{n} \sum_{t=1}^n (\Pi_1 - \widehat{\Pi}_1)' W_{1t} W_{1t}' (\Pi_1 - \widehat{\Pi}_1) \right\| \\ &\leqslant \|\Pi_1 - \widehat{\Pi}_1\|^2 \, \|W_1' W_1 / n\| \\ &= O_P(n^{-1}), \end{aligned}$$

since $\|W_1' W_1 / n\| = O_P(1)$. Under Assumption 12, $\lambda_{K,n}^2 / n = o_P(\sqrt{K/n})$ since B-splines and wavelets satisfy $\lambda_{K,n} \lesssim 1$. Consequently, $\|\widetilde{R}_n' \widetilde{R}_n / n\| = o_P(\sqrt{K/n})$.

Factor $\|\widetilde{B}_\pi' R_n / n\|$ is also straightforward, but depends on sieve dimension K,

$$\begin{aligned} \|\widetilde{B}_\pi' R_n / n\| &\leqslant \left\| \frac{1}{n} \sum_{t=1}^n \widetilde{b}_\pi^K(W_{2t}) (\widehat{\epsilon}_{1t} - \epsilon_{1t}) \right\| \\ &= \left\| \frac{1}{n} \sum_{t=1}^n \widetilde{b}_\pi^K(W_{2t}) W_{1t}' (\Pi_1 - \widehat{\Pi}_1) \right\| \\ &\leqslant \|\Pi_1 - \widehat{\Pi}_1\| \, \|\widetilde{B}_\pi' W_1 / n\| \\ &= O_P(\sqrt{K/n}), \end{aligned}$$

since $\|\widetilde{B}_\pi' W_1 / n\| = O_P(\sqrt{K})$ as the column dimension of W_1 is fixed. The claim then follows by noting $O_P(\sqrt{K/n})$ is the dominating order of convergence. $\square$

Proof of Theorem 3.2. Since $\widehat{\Pi}_1$ the least squares estimator of a linear equation, the rate of convergence is the parametric rate $n^{-1/2}$. The first result is therefore immediate.

For the second step, we consider

$$\|\widehat{\Pi}_2 - \Pi_2\|_\infty \leqslant \|\widehat{\Pi}_2 - \widehat{\Pi}_2^*\|_\infty + \|\widehat{\Pi}_2^* - \Pi_2\|_\infty,$$

and bound explicitly the first right-hand side term. For a given component of the regression function,

$$|\widehat{\pi}_2(x) - \widehat{\pi}_2^*(x)| \leqslant |I| + |II|.$$

We now control each term on the right side.

(1) It holds

$$\begin{aligned}
|I| &\leqslant \|\widetilde{b}_\pi^K(x)\| \, \|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n)^-\| \, \|(\widehat{\widetilde{B}}_\pi - \widetilde{B}_\pi)'Y/n\| \\
&\leqslant \sup_{x\in\mathcal{W}_2} \|\widetilde{b}_\pi^K(x)\| \, \|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n)^-\| \, \|(\widehat{\widetilde{B}}_\pi - \widetilde{B}_\pi)'Y/n\| \\
&\leqslant \zeta_{K,n}\lambda_{K,n} \|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n)^-\| \, \|(\widehat{\widetilde{B}}_\pi - \widetilde{B}_\pi)'Y/n\|.
\end{aligned}$$

Let $\mathcal{A}_n$ denote the event on which $\|\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n - I_K\| \leqslant 1/2$, so that $\|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n)^-\| \leqslant 2$ on $\mathcal{A}_n$. Notice that since $\|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n) - (\widetilde{B}_\pi'\widetilde{B}_\pi/n)\| = o_P(1)$ (Lemma B.2) and, by assumption, $\|\widetilde{B}_\pi'\widetilde{B}_\pi/n - I_K\| = o_P(1)$, then $\mathbb{P}(\mathcal{A}_n^c) = o(1)$. On $\mathcal{A}_n$ then

$$|I| \lesssim \zeta_{K,n}\lambda_{K,n}^2 \|(\widehat{B}_\pi - B_\pi)'Y/n\| = \zeta_{K,n}\lambda_{K,n}^2 \|R_n'Y/n\|.$$

From $R_n'Y = \sum_{t=1}^n b_\pi^K(W_{2t})(\widehat{\epsilon}_{1t} - \epsilon_{1t})Y_t = (\Pi_1 - \widehat{\Pi}_1)'W_1'Y$ it follows that

$$\|R_n'Y/n\| \leqslant \|\Pi_1 - \widehat{\Pi}_1\| \, \|W_1'Y/n\|$$

on $\mathcal{A}_n$, meaning

$$|I| = O_P\left(\zeta_{K,n}\lambda_{K,n}^2/\sqrt{n}\right)$$

as $\|W_1'Y/n\| = O_P(1)$ and $\mathbb{P}(\mathcal{A}_n^c) = o(1)$.

(2) Again we proceed by uniformly bounding II according to

$$|II| \leqslant \zeta_{K,n}\lambda_{K,n} \|(\widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n)^- - (\widetilde{B}_\pi'\widetilde{B}_\pi/n)^-\| \, \|\widetilde{B}_\pi'Y/n\|.$$

The last factor has order $\|\widetilde{B}_\pi'Y/n\| = O_P(\sqrt{K})$ since $\widetilde{B}_\pi$ is growing in row dimension with K. For the middle term, introduce

$$\Delta_B := \widehat{\widetilde{B}}_\pi'\widehat{\widetilde{B}}_\pi/n - \widetilde{B}_\pi'\widetilde{B}_\pi/n$$

and event

$$\mathcal{B}_n := \left\{ \|(\widetilde{B}_\pi' \widetilde{B}_\pi/n)^- \Delta_B\| \leqslant 1/2 \right\} \cap \left\{ \|\widetilde{B}_\pi' \widetilde{B}_\pi/n - I_K\| \leqslant 1/2 \right\}.$$

On $\mathcal{B}_n$, we can apply the bound (Horn and Johnson, 2012)

$$\|(\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi/n)^- - (\widetilde{B}_\pi' \widetilde{B}_\pi/n)^-\| \leqslant \frac{\|(\widetilde{B}_\pi' \widetilde{B}_\pi/n)^-\|^2 \, \|\Delta_B\|}{1 - \|(\widetilde{B}_\pi' \widetilde{B}_\pi/n)^- \Delta_B\|} \lesssim \|\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi/n - \widetilde{B}_\pi' \widetilde{B}_\pi/n\|.$$

Since $\|\widehat{\widetilde{B}}_\pi' \widehat{\widetilde{B}}_\pi/n - \widetilde{B}_\pi' \widetilde{B}_\pi/n\| = O_P(\sqrt{K/n})$ by Lemma B.2, we get

$$|II| = O_P\left(\zeta_{K,n} \lambda_{K,n} \frac{K}{\sqrt{n}}\right)$$

on $\mathcal{B}_n$. Finally, using $\mathbb{P}((A \cap B)^c) \leqslant \mathbb{P}(A^c) + \mathbb{P}(B^c)$ we note that $\mathbb{P}(\mathcal{B}_n^c) = o(1)$ so that the bound asymptotically holds irrespective of event $\mathcal{B}_n$.

Thus, we have shown that

$$\begin{aligned} |\widehat{\pi}_2(x) - \widehat{\pi}_2^*(x)| &\leqslant O_P\left(\zeta_{K,n} \lambda_{K,n}^2 \frac{1}{\sqrt{n}}\right) + O_P\left(\zeta_{K,n} \lambda_{K,n} \frac{K}{\sqrt{n}}\right) \\ &= O_P\left(\zeta_{K,n} \lambda_{K,n} \frac{K}{\sqrt{n}}\right) \end{aligned}$$

as clearly $\sqrt{n}^{-1} = o(K/\sqrt{n})$ and, as discussed in the proof of Lemma B.2, $\lambda_{K,n}^2/n = o_P(\sqrt{K/n})$. This bound is uniform in x and holds for each of the (finite number of) components of $\widehat{\Pi}_2$, therefore the proof is complete. □

B.4 Theorem 4.1

Before proving impulse response consistency, I show that compositions of the model's autoregressive nonlinear maps are also consistently estimated at any fixed horizon. This means that the "functional moving average" coefficient matrices Γ_j involved in Proposition 4.1 can be consistently estimated with $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$.

Lemma B.3. *Under the assumptions of Theorem 3.2 and for any fixed integer $j \geqslant 0$ it holds*

$$\|\widehat{\Gamma}_j - \Gamma_j\|_\infty = o_P(1).$$

Proof. By definition, recall that $\Gamma(L) = \Psi(L)G(L)$ where $\Psi = (I_d - A(L)L)^{-1}$. Since $\Psi(L)$ is an MA(∞) lag polynomial, we have that

$$\Gamma(L) = \left(\sum_{k=0}^{\infty} \Psi_k L^k\right) (G_0 + G_1 L + \ldots + G_p L^p),$$

where $\Psi_0 = I_d$, $\{\Psi_k\}_{k=1}^\infty$ are purely real matrices and G_0 is a functional vector that may also contain linear components (i.e. allow linear functions of X_t). This means that Γ_j is a convolution of real and functional matrices,

$$\Gamma_j = \sum_{k=1}^{\min\{j,p\}} \Psi_{j-k} G_k.$$

The linear coefficients of $A(L)$ can be consistently estimated by $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$, and thus plug-in estimate $\widehat{\Psi}_j$ is consistent for Ψ_j (Lütkepohl, 2005). Therefore,

$$\begin{aligned}
\|\widehat{\Gamma}_j - \Gamma_j\|_\infty &\leqslant \sum_{k=1}^{\min\{j,p\}} \left\|\Psi_{j-k}G_k - \widehat{\Psi}_{j-k}\widehat{G}_k\right\|_\infty \\
&\leqslant \sum_{k=1}^{\min\{j,p\}} \left\|\Psi_{j-k} - \widehat{\Psi}_{j-k}\right\|_\infty \|G_k\|_\infty + \left\|\widehat{\Psi}_{j-k}\right\|_\infty \left\|G_k - \widehat{G}_k\right\|_\infty \\
&\leqslant \sum_{k=1}^{\min\{j,p\}} o_p(1)C_{G,k} + O_P(1)o_p(1) \\
&= o_p(1),
\end{aligned}$$

where $C_{G,k}$ is a constant and $\|G_k - \widehat{G}_k\|_\infty = o_p(1)$ as a direct consequence of Proposition 3.2. □

Note. Since we assume that the model respects either contractivity or stability conditions, the impulse responses must decay (eventually) exponentially fast to zero. This means that by "stitching" bounds appropriately, one should also be able to achieve convergence *uniformly* over $h = 0, 1, \ldots, \infty$.

Recall now that the sample estimate for the relaxed-shock impulse response is

$$\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) = \Theta_{h,\cdot 1}\delta\, n^{-1} \sum_{t=1}^{n} \rho(\widehat{\epsilon}_{1t}) + \sum_{j=0}^{h} \widehat{V}_{j,\ell}(\delta)$$

where

$$\widehat{V}_{j,\ell}(\delta) = \frac{1}{n-j} \sum_{t=1}^{n-j} \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widehat{\widetilde{\delta}}_t\big) = \frac{1}{n-j} \sum_{t=1}^{n-j} \left[\widehat{\Gamma}_j \widehat{\gamma}_j(X_{t+j:t}; \widehat{\widetilde{\delta}}_t) - \widehat{\Gamma}_j X_{t+j}\right].$$

Therefore, the estimated horizon h impulse response of the ℓth variable is

$$\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) := \widehat{\Theta}_{h,\ell 1}\delta\, n^{-1} \sum_{t=1}^{n} \rho(\widehat{\epsilon}_{1t}) + \sum_{j=0}^{h} \left[\frac{1}{n-j} \sum_{t=1}^{n-j} \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widehat{\widetilde{\delta}}_t\big)\right].$$

Lemma B.4. *Under the assumptions of Theorem 4.1 , let $x_{j:0} = (x_j, \ldots, x_0) \in \mathcal{X}^j$ and $\varepsilon \in \mathcal{E}_1$ be nonrandom quantities. Let $\widetilde{\delta}$ be the relaxed shock determined by δ, ρ and ε. Then*

(i) $\sup_{x_{j:0},\varepsilon} |\widehat{\gamma}_j(x_{j:0}; \widetilde{\delta}) - \gamma_j(x_{j:0}; \widetilde{\delta})| = o_P(1)$,

(ii) $\sup_{x_{j:0},\varepsilon} |\widehat{v}_{j,\ell}\big(x_{j:0}; \widetilde{\delta}\big) - v_{j,\ell}\big(x_{j:0}; \widetilde{\delta}\big)| = o_P(1)$,

for any fixed integers $j \geqslant 0$ and $\ell \in \{1, \ldots, d\}$.

Proof.

(i) From Proposition 4.1, we have that

$$\widehat{\gamma}_j(x_{j:0};\delta) = x_j + \Theta_{j,11}\delta\rho(\varepsilon) + \sum_{k=1}^{j}(\Gamma_{k,11}x_{j-k}(\widetilde{\delta}) - \Gamma_{k,11}x_{j-k}),$$

thus

$$\begin{aligned}|\widehat{\gamma}_j(x_{j:0};\delta) - \gamma_j(x_{j:0};\delta)| &= \left|\sum_{k=1}^{j}\left[(\widehat{\Gamma}_{k,11}x_{j-k}(\widetilde{\delta}) - \widehat{\Gamma}_{k,11}x_{j-k}) - (\Gamma_{k,11}x_{j-k}(\widetilde{\delta}) - \Gamma_{k,11}x_{j-k})\right]\right| \\ &\leqslant \sum_{k=1}^{j}\left|\widehat{\Gamma}_{k,11}x_{j-k}(\widetilde{\delta}) - \Gamma_{k,11}x_{j-k}(\widetilde{\delta})\right| + \sum_{k=1}^{j}\left|\widehat{\Gamma}_{k,11}x_{j-k} - \Gamma_{k,11}x_{j-k}\right|.\end{aligned}$$

This yields

$$\sup_{x_{j:0},\varepsilon}|\widehat{\gamma}_j(x_{j:0};\widetilde{\delta}) - \gamma_j(x_{j:0};\widetilde{\delta})| \leqslant 2j\sup_{x\in\mathcal{X}}\left|\widehat{\Gamma}_{k,11}x - \Gamma_{k,11}x\right|.$$

Since j is finite and fixed and the uniform consistency bound of Lemma B.3 holds, a fortiori $\sup_{x\in\mathcal{X}}\left|\widehat{\Gamma}_{k,11}x - \Gamma_{k,11}x\right| = o_P(1)$.

(ii) Similarly to above,

$$\begin{aligned}|\widehat{v}_{j,\ell}\big(x_{j:0};\widetilde{\delta}\big) - v_{j,\ell}\big(x_{j:0};\widetilde{\delta}\big)| &= \left|\Big(\widehat{\Gamma}_{j,\ell}\widehat{\gamma}_j(x_{j:0};\widetilde{\delta}) - \Gamma_{j,\ell}\gamma_j(x_{j:0};\widetilde{\delta})\Big) - \Big(\widehat{\Gamma}_{j,\ell}x_j - \Gamma_{j,\ell}x_j\Big)\right| \\ &\leqslant \|\widehat{\Gamma}_{j,\ell} - \Gamma_{j,\ell}\|_\infty + \|\Gamma_{j,\ell}\|_\infty|\widehat{\gamma}_j(x_{j:0};\delta) - \gamma_j(x_{j:0};\delta)| \\ &\quad + |\widehat{\Gamma}_{j,\ell}x_j - \Gamma_{j,\ell}x_j| \\ &\leqslant 2\|\widehat{\Gamma}_{j,\ell} - \Gamma_{j,\ell}\|_\infty + C_{\Gamma,j,l}\,|\widehat{\gamma}_j(x_{j:0};\delta) - \gamma_j(x_{j:0};\delta)|,\end{aligned}$$

where we have used that $\gamma_j(x_{j:0};\widetilde{\delta}) \in \mathcal{X}$ to derive the first term in the second line. In the last line, $C_{\Gamma,j,l}$ is a constant such that

$$\|\Gamma_{j,\ell}\|_\infty \leqslant \sum_{k=1}^{\min\{j,p\}}\|\Psi_{j-k}\|_\infty\|G_k\|_\infty \leqslant C_{\Gamma,j,l}.$$

The claim then follows thanks to Lemma B.3 and (i).

□

In what follows, define $\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big)$ to be a version of $v_{j,\ell}$ that is constructed using coefficient estimates from $\{\widehat{\Pi}_1,\widehat{\Pi}_2\}$ but evaluated on the true innovations ϵ_t.

Proof of Theorem 4.1. If we introduce

$$\widetilde{\mathrm{IRF}}_{h,\ell}(\delta)^* := \widehat{\Theta}_{h,\ell 1}\delta\, n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t}) + \sum_{j=0}^{h}\left[\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big)\right],$$

then clearly

$$\begin{aligned}\left|\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) - \widetilde{\mathrm{IRF}}_{h,\ell}(\delta)\right| &\leqslant \left|\widehat{\widetilde{\mathrm{IRF}}}_{h,\ell}(\delta) - \widetilde{\mathrm{IRF}}^*_{h,\ell}(\delta)\right| + \left|\widetilde{\mathrm{IRF}}^*_{h,\ell}(\delta) - \widetilde{\mathrm{IRF}}_{h,\ell}(\delta)\right| \\ &= I + II.\end{aligned}$$

To control II, we can observe

$$\begin{aligned} II &\leqslant \left|\widehat{\Theta}_{h,\ell 1}\delta\, n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t}) - \Theta_{h,\ell 1}\delta\mathbb{E}[\rho(\epsilon_{1t})]\right| \\ &\quad + \sum_{j=0}^{h}\left|\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - \mathbb{E}[v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}\big)]\right| \\ &\leqslant \delta\left|\widehat{\Theta}_{h,\ell 1} - \Theta_{h,\ell 1}\right|\left|n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t})\right| + \delta\left|\widehat{\Theta}_{h,\ell 1}\right|\left|n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t}) - \mathbb{E}[\rho(\epsilon_{1t})]\right| \\ &\quad + \sum_{j=0}^{h}\left|\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - \mathbb{E}[v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}\big)]\right| \\ &\leqslant \delta\left|\widehat{\Theta}_{h,\ell 1} - \Theta_{h,\ell 1}\right|\left|n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t})\right| + \delta\left|\widehat{\Theta}_{h,\ell 1}\right|\left|n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t}) - \mathbb{E}[\rho(\epsilon_{1t})]\right| \\ &\quad + \sum_{j=0}^{h}\left|\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big)\right| \\ &\quad + \sum_{j=0}^{h}\left|\frac{1}{n-j}\sum_{t=1}^{n-j}v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - \mathbb{E}[v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}\big)]\right|.\end{aligned}$$

The first two terms in the last bound are $o_P(1)$ since $\left|\widehat{\Theta}_{h,\ell 1} - \Theta_{h,\ell 1}\right| = o_P(1)$, as discussed in Lemma B.3, and $n^{-1}\sum_{t=1}^{n}\rho(\epsilon_{1t}) \xrightarrow{p} \mathbb{E}[\rho(\epsilon_{1t})]$ by a WLLN. For the other terms in the last sum above, we similarly note that

$$\left|\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big)\right| = o_P(1)$$

from Lemma B.4, while thanks again to a WLLN it holds

$$\left|\frac{1}{n-j}\sum_{t=1}^{n-j}v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big) - \mathbb{E}[v_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}\big)]\right| = o_P(1).$$

Since h is fixed finite, this implies that $II = o_P(1)$.

Considering now I, we can write

$$\begin{aligned} I &\leqslant \delta\left|\widehat{\Theta}_{h,\ell 1}\right|\left|n^{-1}\sum_{t=1}^{n}\rho(\widehat{\epsilon}_{1t}) - \rho(\epsilon_{1t})\right| + \sum_{j=0}^{h}\left|\frac{1}{n-j}\sum_{t=1}^{n-j}\widehat{v}_{j,\ell}\big(X_{t+j:t};\widehat{\widetilde{\delta}}_t\big) - \widehat{v}_{j,\ell}\big(X_{t+j:t};\widetilde{\delta}_t\big)\right| \\ &= I' + I''.\end{aligned}$$

Since by assumption ρ is a bump function, thus continuously differentiable over the range of ϵ_t, by the mean value theorem

$$\left| n^{-1} \sum_{t=1}^{n} \rho(\widehat{\epsilon}_{1t}) - \rho(\epsilon_{1t}) \right| \leqslant n^{-1} \sum_{t=1}^{n} |\rho'_t| \, |\widehat{\epsilon}_{1t} - \epsilon_{1t}|$$

for a sequence $\{\rho'_t\}_{t=1}^n$ of evaluations of first-order derivative ρ' at values $\overline{\epsilon_t}$ in the interval with endpoint ϵ_t and $\widehat{\epsilon}_t$. One can use $|\rho'_t| \leqslant C_{\rho'}$ with a finite positive constant $C_{\rho'}$, and by recalling that $\widehat{\epsilon}_{1t} - \epsilon_{1t} = (\Pi_1 - \widehat{\Pi}_1)' W_{1t}$ one thus gets

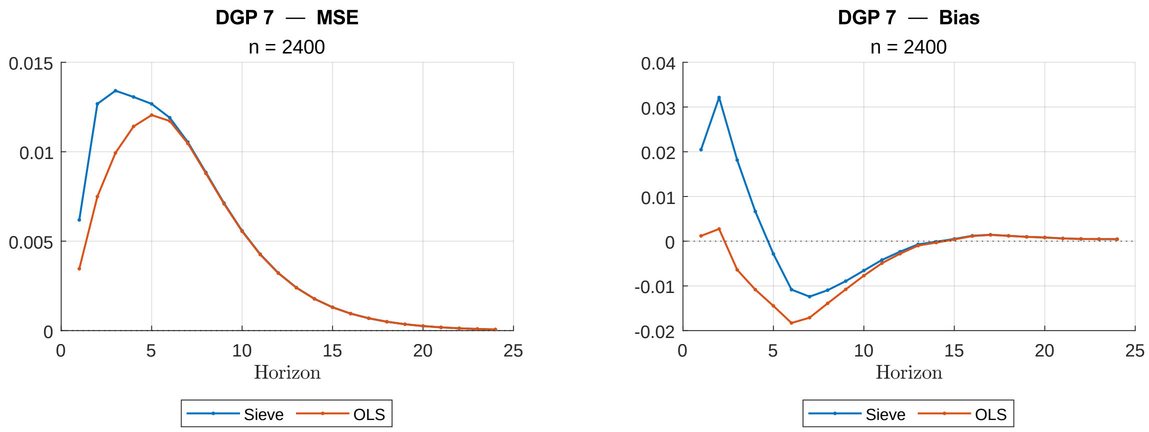
$$\left| n^{-1} \sum_{t=1}^{n} \rho(\widehat{\epsilon}_{1t}) - \rho(\epsilon_{1t}) \right| \leqslant C_{\rho'} \frac{1}{n} \sum_{t=1}^{n} \left| (\Pi_1 - \widehat{\Pi}_1)' W_{1t} \right| \leqslant C_{\rho'} \|\Pi_1 - \widehat{\Pi}_1\|_2 \, \frac{1}{n} \sum_{t=1}^{n} \|W_{1t}\|_2 = o_P(1).$$

This proves that term I' is itself $o_P(1)$. Finally, to control I'', we use that by construction estimator $\widehat{\Pi}_2$ is composed of sufficiently regular functional elements i.e. B-spline estimates of order 1 or greater. Thanks again to the mean value theorem

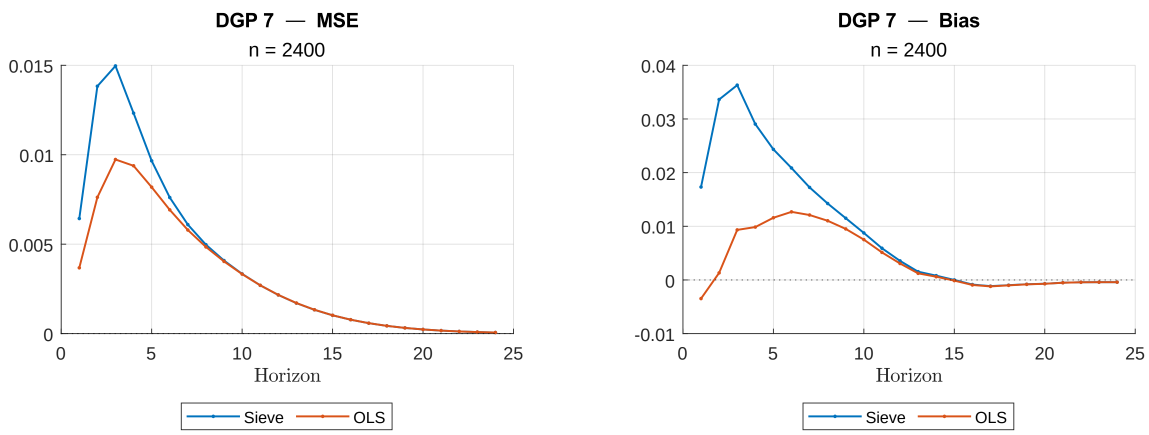
$$\begin{aligned} \left| \frac{1}{n-j} \sum_{t=1}^{n-j} \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widehat{\widetilde{\delta}}_t\big) - \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widetilde{\delta}_t\big) \right| &\leqslant \frac{1}{n-j} \sum_{t=1}^{n-j} \left| \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widehat{\widetilde{\delta}}_t\big) - \widehat{v}_{j,\ell}\big(X_{t+j:t}; \widetilde{\delta}_t\big) \right| \\ &\leqslant C_{\widehat{v}',j,\ell} \, \frac{1}{n-j} \sum_{t=1}^{n-j} |\widehat{\epsilon}_{1t} - \epsilon_{1t}| \end{aligned}$$

for any fixed j and some $C_{\widehat{v}',j,\ell} > 0$. This holds since $\widehat{v}_{j,\ell}$ is uniformly continuous by construction. Note that we have assumed that the nonlinear part of Π_2 belongs to a Hölder class with smoothness $s > 1$ (for simplicity, assume here that s is integer, otherwise a similar argument can be made). Then, even though $C_{\widehat{v}',j,\ell}$ depends on the sample, it is bounded above in probability for n sufficiently large. Following the discussion of term I', we deduce that the last line in the display above is $o_p(1)$. As h is finite and independent of n, it follows that also I'' is of order $o_P(1)$. $\square$

C Additional Plots



(a) $\delta = +2$



(b) $\delta = -2$

Figure 8: Simulation results for DGP $2'$ when considering $\tilde{\varphi}$ in place of φ.

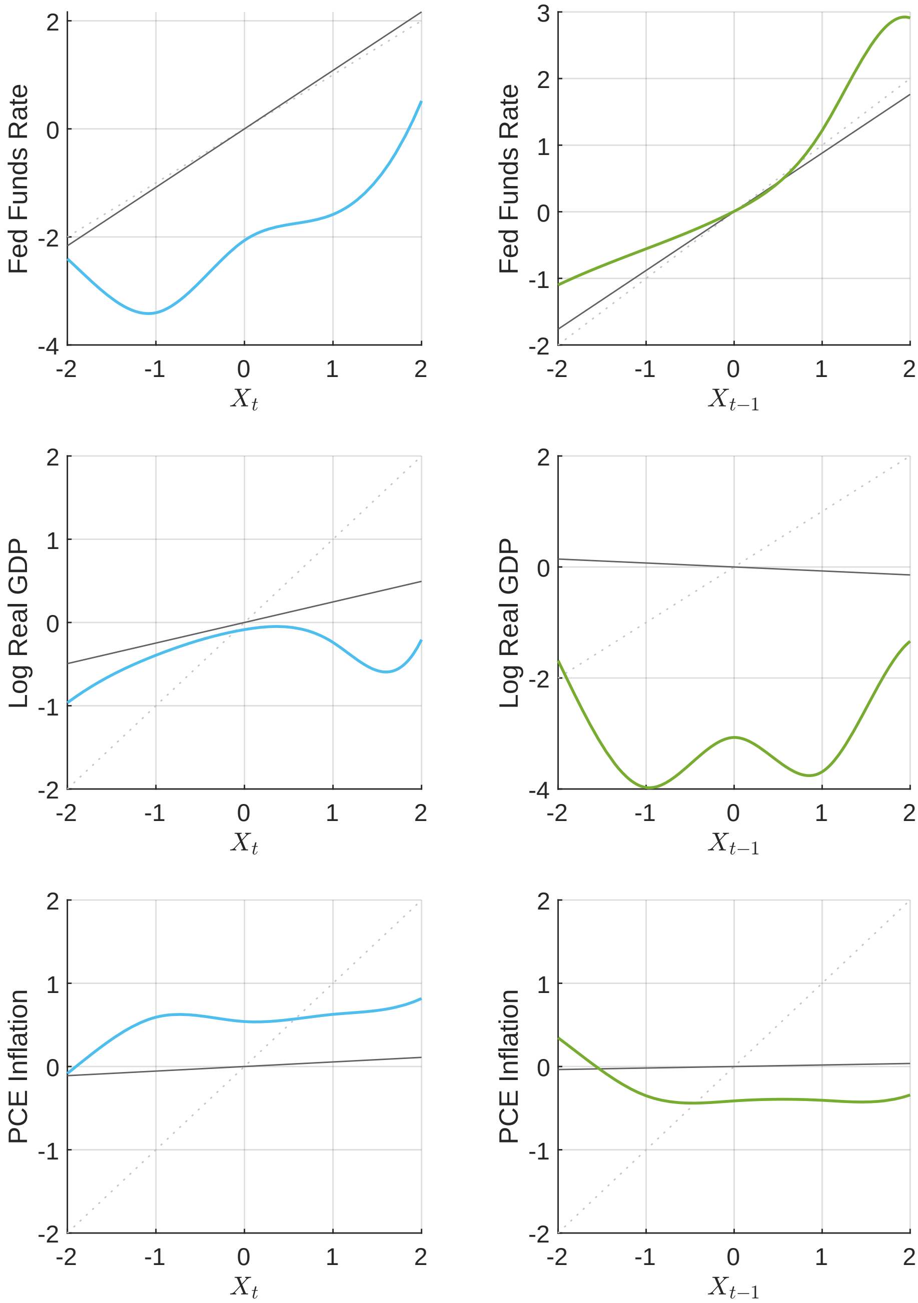
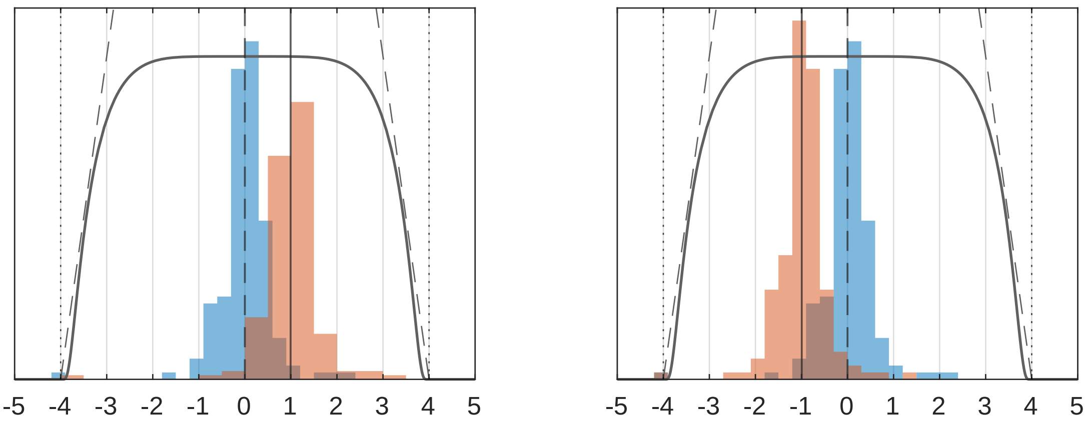
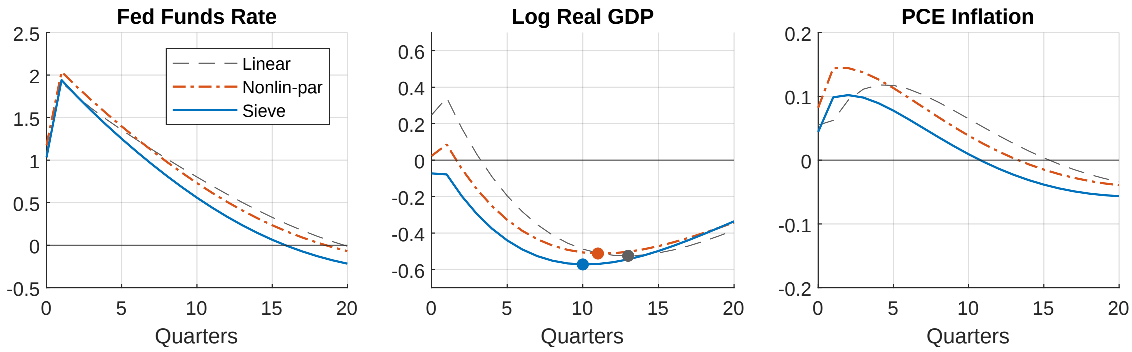


Figure 9: Estimated nonlinear regression functions for the narrative U.S. monetary policy variable. Contemporaneous (left side) and one-period lag (right side) effects are shown, linear and nonlinear functions. For comparison, linear VAR coefficients (dark gray) and the identity map (light gray, dashed) are shown as lines.

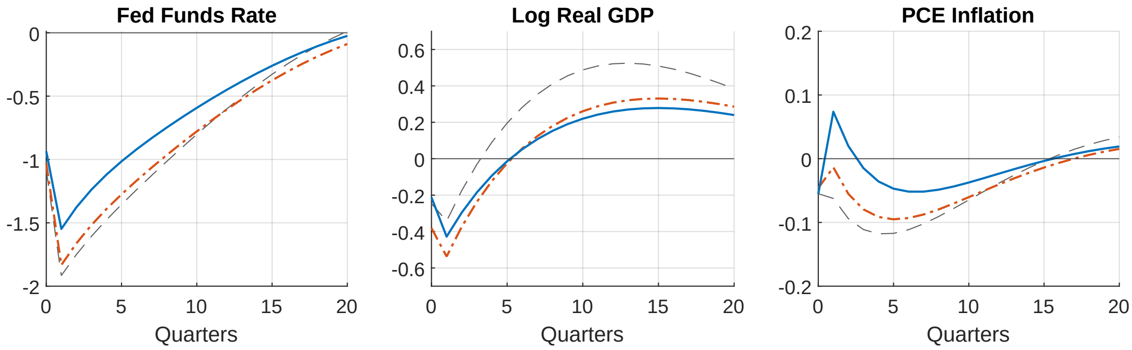


(a) $\delta = +1$ (b) $\delta = -1$

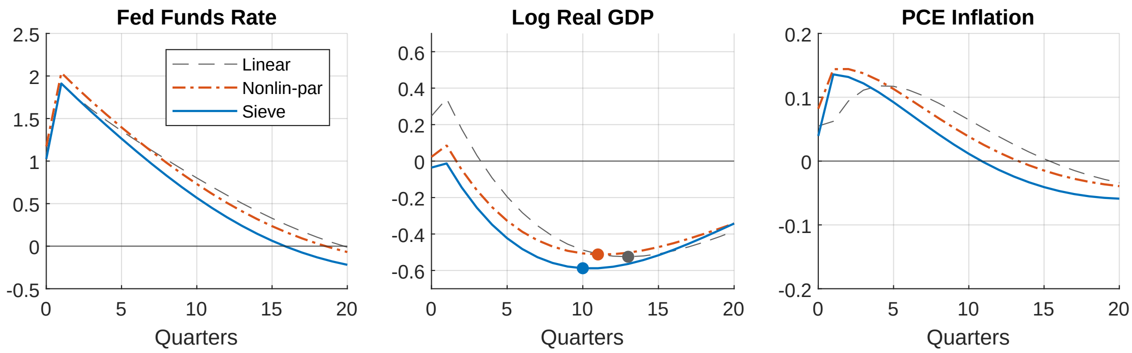
Figure 10: Comparison of histograms and shock relaxation function for a positive (left) and negative (right) shock in monetary policy. Original (blue) versus shocked (orange) distribution of the sample realization of ϵ_{1t}. The dashed vertical line is the mean of the original distribution, while the solid vertical line is the mean after the shock.



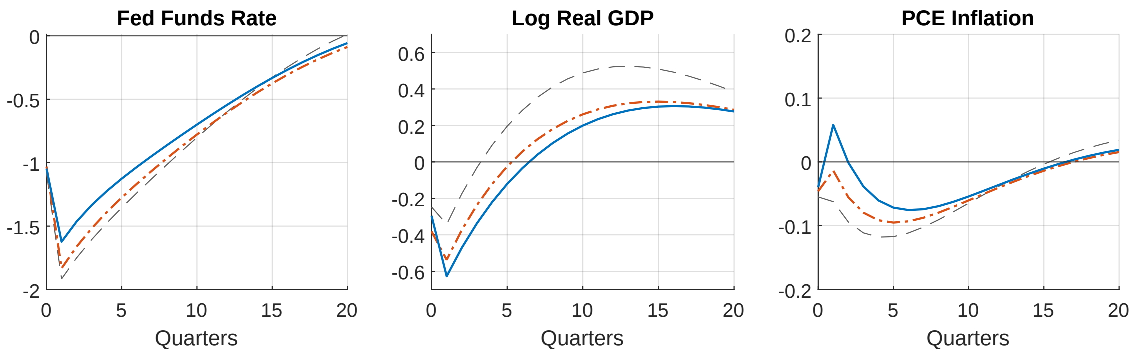
(a) $\delta = +1$, knots at $\{-1, 1\}$



(b) $\delta = -1$, knots at $\{-1, 1\}$



(c) $\delta = +1$, knot at $\{0\}$



(d) $\delta = -1$, knot at $\{0\}$

Figure 11: Robustness plots for U.S. monetary policy shock when changing knots compared to those used in Figure 6. Note that linear and parametric nonlinear responses do not change.

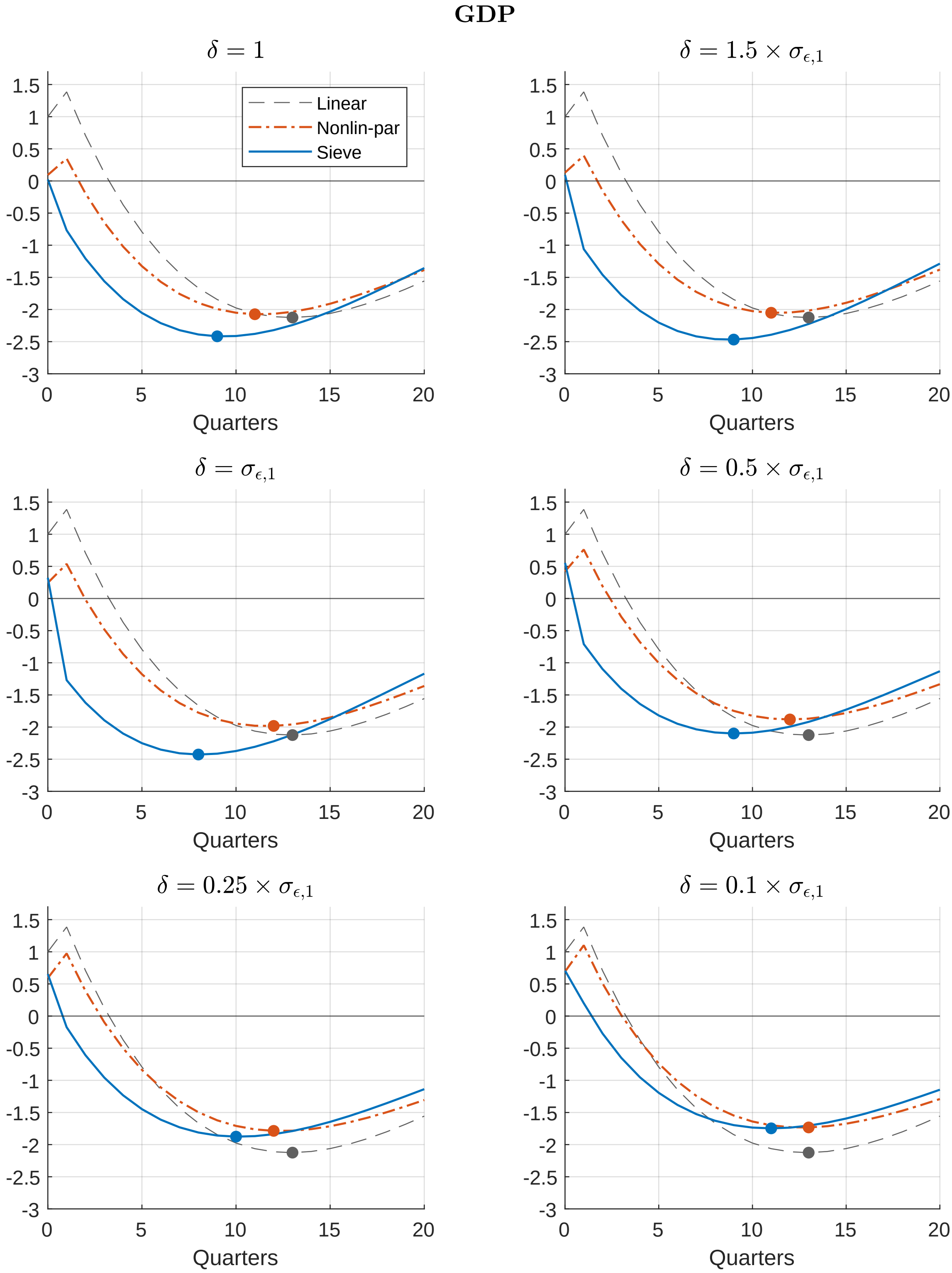


Figure 12: Relative changes in the GDP impulse responses function when the size of the shock is reduced from that used in Figure 6. The standard deviation of $X_t \equiv \epsilon_{1t}$ is $\sigma_{\epsilon,1} \approx 0.5972$. Linear IRFs are re-scaled such that for all values of δ the linear response at $h = 0$ is one in absolute value. Nonlinear IRFs are re-scaled by δ times the linear response scaling factor.

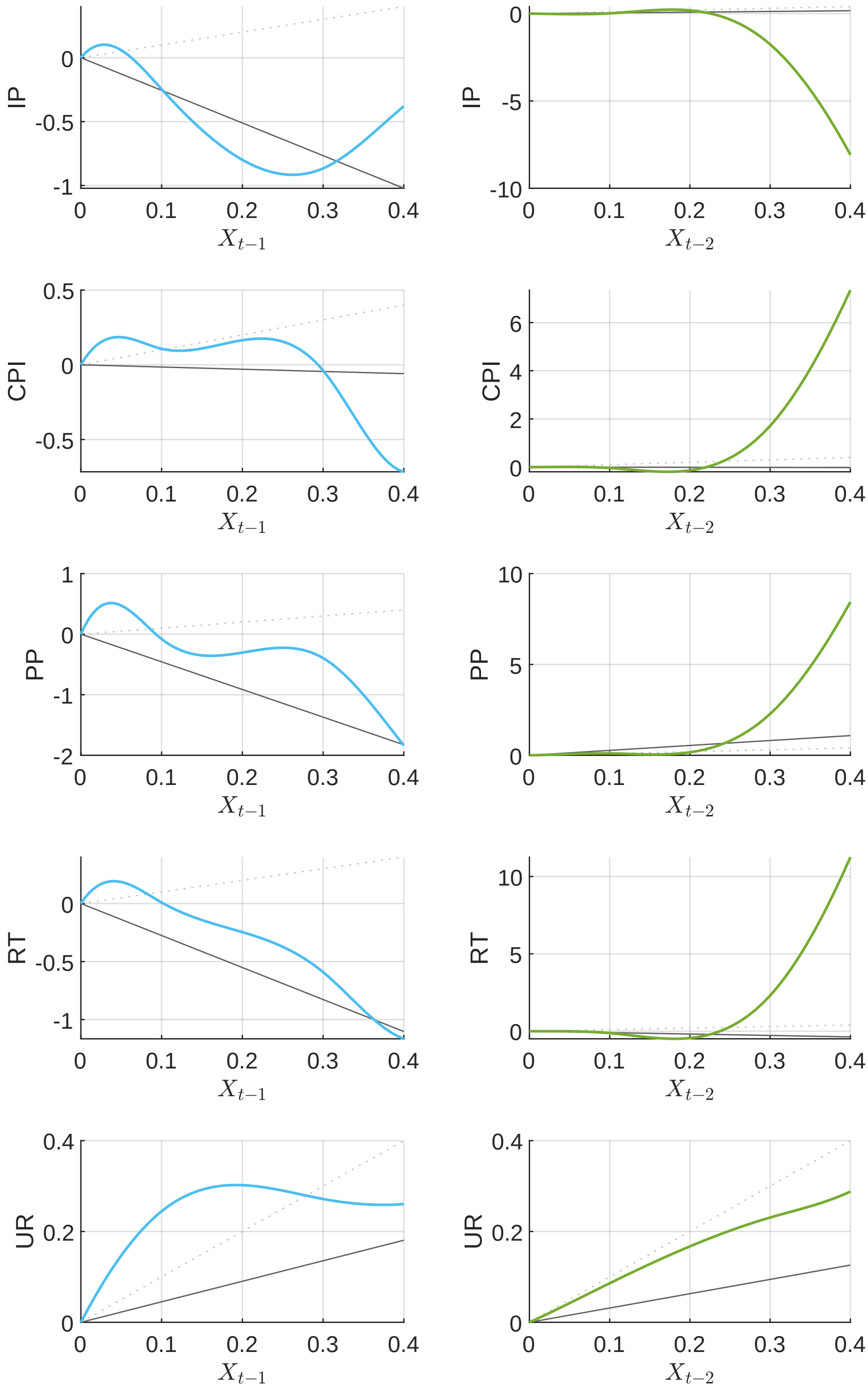
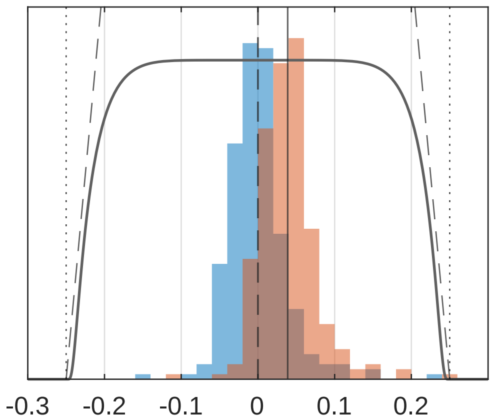
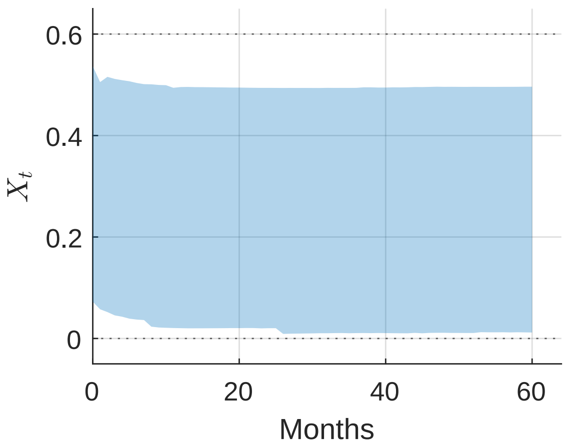


Figure 13: Estimated nonlinear regression functions for the 3M3M subjective interest rate uncertainty measure. One-period (left side) and two-period lag (right side) effects are shown, combining linear and nonlinear functions. For comparison, linear VAR coefficients (dark gray) and the identity map (light gray, dashed) are shown as lines.



(a) $\delta = \sigma_\epsilon$



(b) Envelope

Figure 14: **[Top]** Histograms and shock relaxation function for a one-standard-deviation shock in interest rate uncertainty. Original (blue) versus shocked (orange) distribution of the sample realization of ϵ_{1t}. The dashed vertical line is the mean of the original distribution, while the solid vertical line is the mean after the shock. **[Bottom]** Envelope (min-max) of shocked paths for one-standard-deviation impulse response.

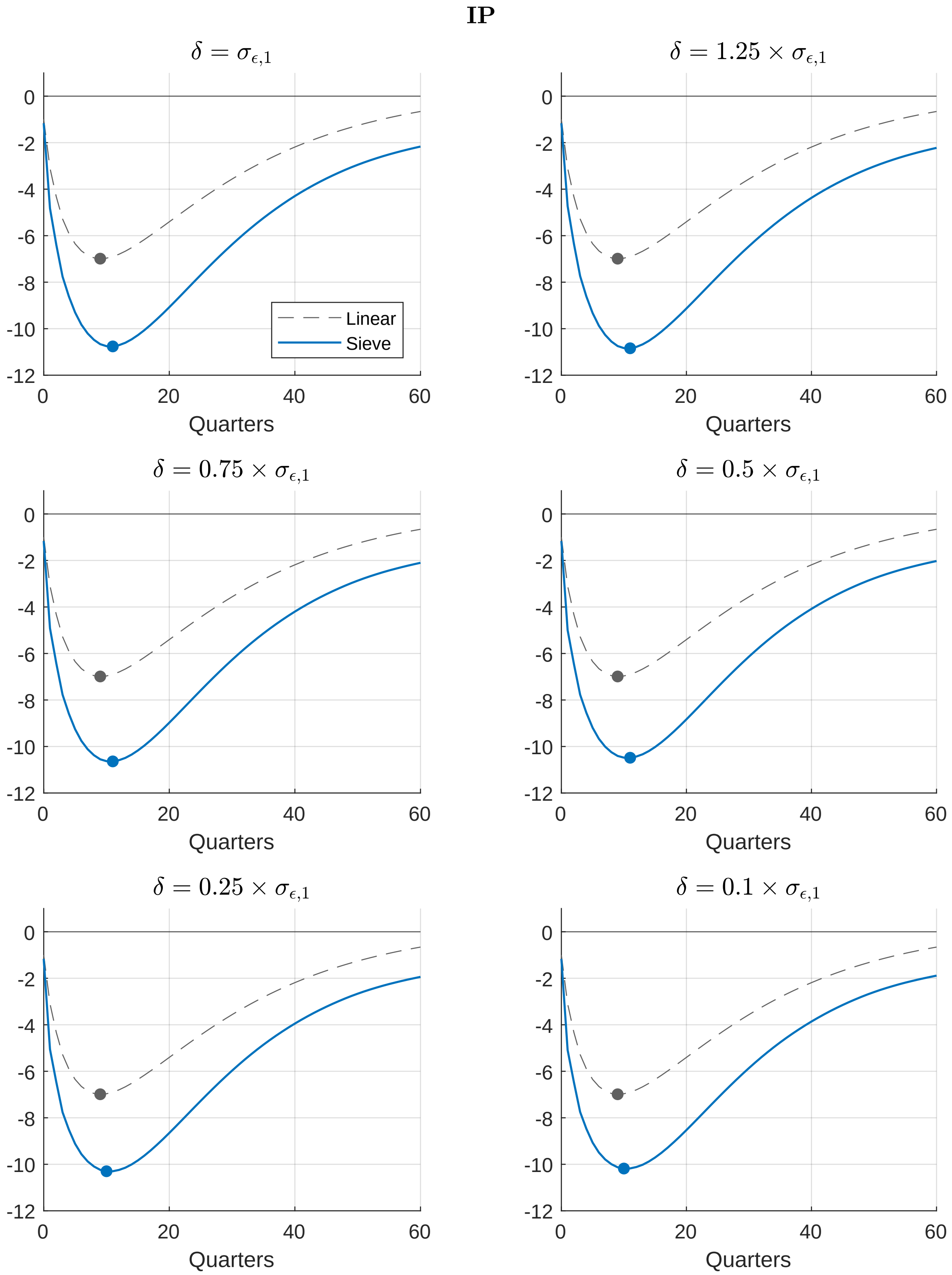


Figure 15: Relative changes in the industrial production impulse responses function when the size of the shock is reduced from that used in Figure 7. The standard deviation of $\equiv \epsilon_{1t}$ is $\sigma_{\epsilon,1} \approx 0.0389$. Linear IRFs are re-scaled such that for all values of δ the linear response at $h = 0$ is one in absolute value. Nonlinear IRFs are re-scaled by δ times the linear response scaling factor.

CPI

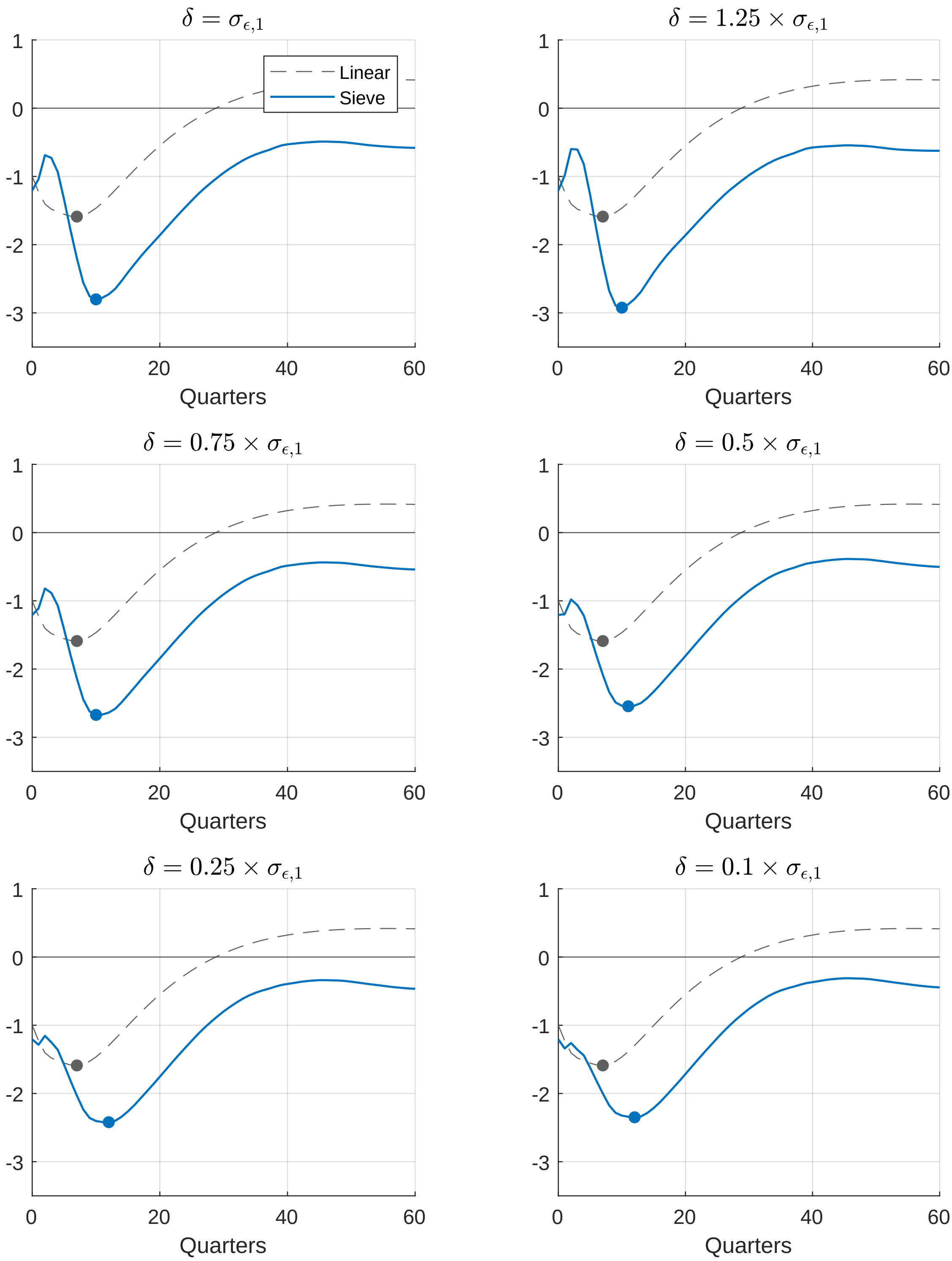


Figure 16: Relative changes in the CPI impulse responses function when the size of the shock is reduced from that used in Figure 7. The standard deviation of $\equiv \epsilon_{1t}$ is $\sigma_{\epsilon,1} \approx 0.0389$. Linear IRFs are re-scaled such that for all values of δ the linear response at $h = 0$ is one in absolute value. Nonlinear IRFs are re-scaled by δ times the linear response scaling factor.