Essays in Time Series Econometrics

Jairo Flores

TESI DOCTORAL UPF / year 2022

THESIS SUPERVISOR
Geert Mesters
Department of Economics and Business

To my beloved Maite and my parents, Sara and Edgar

Acknowledgments

I am incredibly thankful to the numerous talented and gracious individuals who contributed to the success of this dissertation. I would like to thank my advisor, Geert Mesters, for his excellent support, guidance, and patience over the past five years. He believed in my work and my potential and offered close support even during those times when I was not very optimistic. Without his guidance, this dissertation would not have been possible.

I also would like to thank Christian Brownlees, Barbara Rossi, Piotr Zwiernik, Luca Fornaro, Isaac Baley and faculty members and Ph.D. students of UPF for their helpful comments during seminars and presentations, which helped me to improve my work. Marta Araque and Laura Agusti provided fantastic support during the Ph.D.

I am deeply grateful to my beloved Maite for her incredible support and understanding over the past years. I am lucky to have her as my partner. I also thank my parents, Sara and Edgar, my sister Dariana, and my nephew Gael; their presence keeps me grounded and reminds me of my priorities. I am also grateful to my extended family, Rosa, Luis, Juan, Ena, Hugo, Nancy, Dalila, Cesar, Ana, Victoria, and Arnulfo, for their constant support and advice. Special mention for those who departed soon: Juan, Victor, and Pablo.

I thank my forever friends and closest colleagues in the Ph.D. Cristiano Mantovani, Claudia Meza-Cuadra, Hector Paredes, Elisa Failache, András Jagadits, Maria Ptashkina, Barna Szabó, Anna May, Diego Bohorquez, Marco Ceron, Laurenz Baertsch, Carolina Castañeda, Fatima Antelo, and Pablo Carballo. I am thankful for their presence and the happiness they bring into my life. By far, they were the best part of my Ph.D. studies.

Finally, I am also grateful to the Central Reserve Bank of Peru for their financial support and for believing in my potential. I would also like to thank my colleagues and friends Mauricio, Luigui, Jose, Marlon, Jesus, Gladys, Teresa, and Adrian for their patience and support these past years.

Jairo Flores Audante 2022

Abstract

This doctoral thesis develops new methods for time series analysis and applies these to prominent empirical problems in macroeconomics. It contains three chapters. Chapter one proposes a new non-parametric generalized method of moments estimator for handling time-varying parameters. Chapter two proposes a simple and practical method for estimating impulse response functions based on using a single flexible and interpretable function to approximate the impulse response. Chapter three discusses a large-scale simulation study to compare impulse response estimates obtained from vector autoregressive model average models with vector autoregressive and local projection methods.

Resum

Aquesta tesi doctoral desenvolupa nous mètodes per a l'anàlisi de sèries temporals i els aplica a problemes empírics destacats en macroeconomia. Conté tres capítols. El primer capítol proposa un nou mètode generalitzat no paramètric d'estimador de moments per al maneig de paràmetres variables en el temps. El capítol dos proposa un mètode senzill i pràctic per estimar les funcions de resposta a l'impuls basat en l'ús d'una única funció flexible i interpretable per aproximar la resposta a l'impuls. El capítol tres analitza un estudi de simulació a gran escala per comparar les estimacions de resposta a impulsos obtingudes a partir de models mitjans de models autoregressius vectorials amb mètodes autoregressius vectorials i de projecció local.

Contents

List of Figures

List of Tables

Introduction

This doctoral thesis develops new methods for time series analysis and applies these to prominent empirical problems in macroeconomics. It contains three chapters. Chapter one proposes a new non-parametric generalized method of moments (GMM) estimator for handling time-varying parameters. Chapter two proposes a simple and practical method for estimating impulse response functions based on using a single flexible and interpretable function to approximate the impulse response. Chapter three discusses a large-scale simulation study to compare impulse response estimates obtained from vector autoregressive model average models with vector autoregressive and local projection methods.

In the first chapter, I extend the standard GMM framework of Hansen (1982) to allow for time-varying parameters. To conduct inference, I propose a new non-parametric estimator that is based on a local polynomial approximation where the true parameters are locally approximated by a polynomial of specified order. I argue that this method is appealing in cases where there is no information available about the dynamics of the parameters to be estimated, e.g., no random walk or ARMA dynamics are appropriate. As many well-known models can be defined in terms of moments, e.g., moving average and autoregressive conditional heteroskedasticity models, my approach allows extending all these models to allow for time-varying parameters. I develop the asymptotic theory of the proposed local polynomial estimator under the assumption of local stationarity and show that the estimator is consistent and asymptotically normal. The small sample properties of the nonparametric GMM estimator are studied through Monte Carlo simulations for two examples: the

MA(1) model and the ARCH(1) model. In the empirical part of the study, the proposed estimator is applied to a gravity model of international trade with varying parameters and finds that the average patterns are consistent with previous studies but provide new information for the dynamics of distance. Specifically, I find evidence of distance instability of the parameters of the gravity model for the US economy.

In the second chapter, I provide tools for characterizing impulse response functions using their canonical characteristics (peak, location, and half-lives before and after the peak) by introducing functional local projections as a way to summarize the evidence from any impulse response estimator. Honest confidence bands are constructed by taking into account both the bias and variance of the functional approximation. I also propose an extension that allows for time-varying parameters and derive the asymptotic properties of the proposed estimator. In a simulation study, the performance of functional and traditional local projection estimates are compared in terms of their coverage ratios and lengths for different levels of smoothness, sample sizes, and experiments. The results show that functional local projections perform best in terms of coverage when the degree of smoothness is high and worst when it is low, and their performance improves as the sample size increases. In the empirical study, the functional local projection approach to estimating impulse response functions yields smoother and more informative results than the traditional method and provides additional information about the impulse response functions through the estimated parameters of the Gaussian basis function. It also reveals that the impulse response functions may be time-varying, with differences in peaks and timing of peaks at different points in the sample. Overall, the functional local projection method provides valuable practical knowledge into the effects of shocks on the economy that helps to characterize and summarize the impulse response functions.

In the third chapter, I conduct a large-scale simulation study to compare impulse response estimates from structural vector autoregressive moving average (SVARMA) models to those obtained from the more common structural vector autoregressive (SVAR) and local projection models. The evaluation of the impulse response estimates is based on quantifying the

bias and variance across Monte Carlo replications. The simulation design is based on randomly selected dynamic generalized processes from a dynamic factor model known to describe US macroeconomic time series accurately. The SVAR, SVARMA, and LP methods are implemented to estimate the structural IRFs under three identification schemes: observed shocks, recursive, and instrumental variables/proxy. The results show that the SVARMA model exhibits a more extreme bias-variance trade-off than the vector autoregressive model at intermediate horizons, meaning it has a lower bias but higher variance. The SVARMA model also faces a similar trade-off between bias and variance as the VAR model when compared with the LP method. These results suggest that the SVARMA model may be a useful estimator in certain contexts where a more extreme trade-off between bias and variance is desired.

Chapter 1

LOCAL POLYNOMIAL ESTIMATION OF TIME-VARYING PARAMETERS IN GMM

Abstract

In this paper, I propose a new non-parametric GMM estimator for conducting inference on time-varying parameters in nonlinear economic models. The estimator is based on a local polynomial approximation of the time-varying parameters following the approach in Fan and Gijbels (1996) and Kristensen and Lee (2019). I show that the proposed estimator retains the properties of consistency and asymptotic normality of the standard GMM under the assumption of uniform locally stationarity introduced by Dahlhaus et al. (1997). In a Monte Carlo study, the proposed estimator exhibits good performance for various models of interest, such as the moving average and ARCH models. Empirically I study a canonical gravity model for international trade that I extend to allow for varying parameters. I apply the model to the US economy and its main trade partners and find evidence of varying effects of importer GDP and distance on US exports.

1.1 Introduction

The Generalized Method of Moments (GMM), as introduced in Hansen (1982), has become an important tool for inference in econometrics. Among others, this modeling framework has found broad appeal in macroeconomics and finance.[1] In particular, two features of the GMM framework are appealing for its application in dynamic macroeconomic models: (i) the method relies on a vector of moment conditions that can often be derived from optimizing behavior of agents and (ii) under reasonable assumptions it allows for the use of lagged variables as instrumental variables in order to capture the typical complex relationships between economic variables. Both of these characteristics can help to handle partially-defined nonlinear dynamic models (e.g. **?**). Hall (2005) provides a textbook treatment for GMM.

When using economic data, empirical evidence has highlighted the importance of taking into account the possibility of time instabilities in the estimated parameters. For example, Stock and Watson (1996) found significant instability in a large fraction of the univariate and bivariate autoregressive models from 76 representative US monthly macroeconomics time series. Further, Ang and Chen (2007) found that the market beta of the value portfolio varies considerably over time when studying the US stocks market. For dealing with time instability in parameters, most of the current literature has focused on either discrete change models (e.g. Hamilton, 1989; Boldea et al., 2019) or parametric smooth change models (e.g. Primiceri, 2005). Both have their limitations. When using economic data, empirical evidence has highlighted the importance of considering the possibility of time instabilities in the estimated parameters. For example, Stock and Watson (1996) found significant instability in a large fraction of the univariate and bivariate autoregressive models from 76 representative US monthly macroeconomics time series. Further, Ang and Chen (2007) found that the market beta of the value portfolio varies considerably over time when studying the US stock market. For dealing with time instability in parameters, most of the current literature has focused on either discrete change models (e.g. Hamilton, 1989; Boldea et al., 2019) or parametric smooth change models (e.g. Primiceri, 2005). Both have their limitations.

First, while discrete structural changes in parameters are appealing for in-

[1] The applications of GMM in economics and finance are broad and diverse. For a comprehensive list of applications, see Table 1.1 in Hall (2005).

terpretability, sometimes abrupt changes are economically implausible. As such smooth change models can be more appealing.[2] While the vast majority of the macro literature adopts parametric, often random walk, specifications for time-varying parameters (e.g. Primiceri, 2005), there exists little rationale or justification for this. A few works have considered time-varying parameters frameworks where the parameters are assumed to be a deterministic and smooth function of time (e.g., Robinson (1989), Orbe et al. (2000), Cai (2007)).

In this study, I extend the GMM framework to allow for time-varying parameters in a nonparametric fashion. Specifically, parameters are allowed to vary over time smoothly as in the polynomial approximation setup developed in Fan and Gijbels (1996). For this extended GMM model, I propose to estimate the true time-varying parameter path using a local polynomial estimator where the true parameter is *locally* approximated by a polynomial of order n (e.g., $n = 0$ local constant, $n = 1$ local linear, etc.). In this way, I bring nonparametric methods into the time-varying parameter GMM model, which offers several advantages. First, little restrictions are imposed on the functional form of the parameters, unlike the parametric cases in which we need to assume a process for them (e.g., random walk or structural break). Second, the nonparametric estimates of a time-varying parameters model can give helpful information about the shape of the coefficients that can be used in subsequent parametric estimation. Third, the estimator is computationally straightforward because there is no substantial difference from a kernel estimation regression.

In general, parametric models for time-varying coefficients might be more efficient if the coefficient functions are correctly specified, but this is hardly the case in practice. For these reasons, various researchers have been considering nonparametric time-varying/functional coefficients in the context of time series regressions models such as Robinson (1989), Fan et al. (1999), Cai et al. (2000),Cai and Li (2008), Chen and Hong (2012), Chen (2015), among others. I argue that this method is appealing in cases where no information about the dynamic of the parameters to be estimated is given. In light of these works, my contribution is to make these methods applicable to the general GMM framework.

This extension covers many specific models of interest, such as time-varying

[2] As Hansen (2001) comments: *"While it may seems unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect"*.

moving average (MA) models and autoregressive conditional heteroskedasticity (ARCH) models. For example, Yan et al. (2020) points out the importance of considering time-varying parameters in the context of multivariate moving average models, while Chen and Hong (2016) and Inoue et al. (2020) highlight evidence of parameter instability in the family of GARCH models. These models have the additional feature that the implied moment conditions are nonlinear in the parameters, a characteristic the GMM framework can easily handle.

I develop the asymptotic theory of the proposed estimator based on the assumption of *local stationarity* as introduced in Dahlhaus et al. (1997). This concept has been adopted and extended in Dahlhaus et al. (2006), Dahlhaus et al. (2019), and Kristensen and Lee (2019). The idea behind this assumption is that a locally stationary process can be approximated by its local stationary approximation on each point of time, such as modified versions of the law of large numbers, and the central limiting theorem can be applied. Based on this assumption, I show that the proposed estimator retains the properties of consistency and asymptotic normality of the standard GMM. As such, these results can be viewed as complementary to those of Kristensen and Lee (2019), who show similar results for time-varying parameter estimates in a likelihood framework. I use the results from Kristensen and Lee (2019) and Dahlhaus et al. (2019) to establish the asymptotic properties of the local polynomial estimator within the GMM framework.

I study the small sample properties of the estimator in a Monte Carlo study for two examples of interest: the MA(1) model and the ARCH(1) model. In order to assess the behavior of the proposed estimator, I consider three different specifications for the time-varying parameters: cosine function, a linear trend with a break function, and square root function[3], while its performance is evaluated by computing the mean absolute deviation error (MADE) and root of the mean squared errors (RMSE) for each case. I find evidence of a good performance of the proposed estimator in terms of MADE, RMSE, and coverage of the confidence intervals.

In the empirical part, I estimate a gravity model of international trade with "distance"-varying parameters for the US economy and main trade partners. The gravity model is an analytical framework for the study of bilateral trade flows

[3]The linear trend with a break is not part of my framework because I assumed that parameters change smoothly over time, but I consider important to evaluate the performance of the estimator under this case.

(see Eaton and Kortum (2002), Anderson and Van Wincoop (2003), Silva and Tenreyro (2006)). In particular, Kalirajan (1999) and Tzouvelekas (2007) show some evidence of varying coefficients in gravity models. In this sense, I study the effects of distance on US exports focusing on the trade-off between the distance and partners' market sizes (GDP). I explore "distance"- varying-parameter gravity model as in Marimoutou et al. (2010) by ordering the data according to increasing spatial distance: the first observation corresponds to the nearest neighbor of the US (Canada) and the last one to the country very far from US (Indonesia) such that the top 100 trade partners are included in the analysis. I argue that under these definitions, the "distance" can be viewed as "time" in my framework, and the nonparametric GMM estimator can be applied. I find evidence of varying effects of partners' GDP and distance on US exports.

The remainder of this paper is organized as follows. In Section 2, I review the main papers directly related to the present research. In Section 3, I describe some examples of what kind of models the proposed method can handle. Section 4 details the general framework and shows the asymptotic theory results. Section 5 displays the main findings of the Monte Carlo study, and Section 6 shows results from the empirical application. Section 7 concludes. Proofs, tables, and plots can be found in the appendix of chapter 1.

1.2 Literature Review

There is a significant body of literature on nonparametric techniques for dealing with time-varying parameters, particularly in linear models. Robinson (1989) and Robinson (1991) introduced a class of kernel estimators for time-varying parameters in linear models. He also demonstrated that making the time-varying parameter depend on the sample size is necessary to provide the asymptotic justification for any nonparametric smoothing estimators. The "intensity" assumption suggests that consistent estimation can be achieved through an increasing density of data points or an "intense" sampling of data points. In a series of papers Orbe et al. (2000), Orbe et al. (2005), and Orbe et al. (2006) built on the work of Robinson (1989) by including seasonal patterns in the analysis.

It is important to note that the estimators discussed above are local constant estimators or Nadaraya-Watson estimators, which are known to have a larger bias and can be affected by boundary effects. In this sense, Cai (2007) proposes a local linear approach to estimate the time trend and coefficient functions and

studies the asymptotic properties of these estimators under certain mixing conditions. Subsequent studies extend the nonparametric local linear estimation of time-varying parameters even further: Chen and Hong (2012) propose a consistent test for smooth structural changes and abrupt structural breaks with known or unknown change points. The idea is to estimate smooth time-varying parameters by local smoothing and compare the fitted values of the restricted constant and unrestricted time-varying parameters. Chen (2015) addresses the problem of modeling and detecting parameter stability in econometric models, focusing on models with endogenous regressors. He proposes a local linear two-stage least squares estimation method to estimate coefficient functions in a time-varying coefficient time series model with potential time-varying endogeneity.

From the parametric point of view, there are different ways to deal with time-varying parameters in the literature. Some papers related to structural breaks are, for example, Hamilton (1989), Bai and Perron (1998), Tsay (1998), Kim et al. (1999), Sims (1999), Sims and Zha (2001) and Sims and Zha (2006), and references therein. This model assumes time-varying parameters to allow for abrupt structural breaks in economic relationships and obtain efficient estimation. Another strand of the literature assumes stochastic coefficients generated by a random walk. Classic examples of this part of the literature are Primiceri (2005), who studies changing monetary policy, and Cogley and Sargent (2005) regarding evolving inflation dynamics.

The present research is also related to the literature on locally stationary processes. Dahlhaus et al. (1997) introduced a general minimum distance estimation procedure for nonstationary time series. Dahlhaus et al. (2006) generalized the ARCH(∞) model to the nonstationary class of ARCH(∞) models with time-varying coefficients, which can be approximated as locally stationary ARCH(∞) processes at fixed time points. Dahlhaus et al. (2019) presented a general theory for locally stationary processes, including their stationary approximation and stationary derivative, as well as laws of large numbers, central limit theorems, and bias expansions.

A paper closely related to the present study is Kristensen and Lee (2019), in which the authors develop a local polynomial (quasi-) maximum likelihood estimator of time-varying parameters under the assumption of locally stationary processes. The main difference is that our framework is based on GMM, leading to new results on the use of polynomial fitting to estimate time-varying parameters.

1.3 Examples

In this section I first introduce some examples that show that the framework under study deals with typical cases of interest in which there are moments conditions which are time-varying and nonlinear in the parameters.

1.3.1 Moving Average Models

A moving average model is a time series model used to forecast future values based on past observations. The model is based on the idea that the future value of a time series can be predicted by taking the average of a certain number of previous innovations.

Let us consider the simple MA(1) model for outcome variables y_t given by

$$y_t = e_t - \beta e_{t-1} \,, \qquad t = 1, \ldots, n \,,$$

where e_t has mean zero, variance σ^2 and is uncorrelated over time. The parameters of this model can be identified from the following moment's conditions

$$\mathbb{E}(y_t^2) = \sigma^2(1+\beta^2) \qquad \text{and} \qquad \mathbb{E}(y_t y_{t-1}) = -\sigma^2\beta \,.$$

Notice that the moment conditions are nonlinear in the model parameters.

The parameters are summarized in the vector $\theta = (\beta, \sigma^2)$. Making these parameters time-varying has been important in various applications; see for examples Triantafyllopoulos and Nason (2007), and Yan et al. (2020). We consider $\theta_t = (\beta_t, \sigma_t^2)$ and write the time-varying moment conditions as

$$\mathbb{E}[g(v_t, \theta_t)] = 0 \qquad \text{with} \qquad g(v_t, \theta_t) = \begin{bmatrix} y_t^2 - \sigma_t^2(1+\beta_t^2) \\ y_t y_{t-1} + \sigma_t^2 \beta_t \end{bmatrix} \,.$$

Here $v_t = (y_t, y_{t-1})'$ and the known function g captures the non-linear and time-varying moment restrictions. In this paper, I will treat the parameters θ_t deterministic. I propose a new estimator and develop the corresponding theory.

1.3.2 ARCH models

The autoregressive conditional heteroskedastic (ARCH) model is a statistical model used to model the time-varying volatilities. It is often used to model and

predict the variance or volatility of returns in financial markets, while its original application was for modeling the variance of inflation series (e.g. Engle, 1982).

We consider a simple ARCH(1) model

$$y_t = \sqrt{\lambda_t} e_t \,, \qquad \lambda_t = w + \alpha y_{t-1}^2 \,, \qquad t = 1, \ldots, n \,,$$

where e_t has mean zero, variance one, and is uncorrelated over time. The parameters of this model $\theta = (\omega, \alpha)'$ are typically estimated by either parametric maximum likelihood methods, i.e., under an additional distributional assumption for e_t, or by using moment methods.

When we adopt moment methods to estimate the parameters in this model, typical choices for the moments include

$$\mathbb{E}(y_t^2(y_t^2 - \lambda_t)) = 0 \qquad \text{and} \qquad \mathbb{E}\left(y_t^2\right) = \frac{\omega}{1-\alpha} \,.$$

Additionally, we can over-identify the model parameters by also including the fourth-moment condition

$$\mathbb{E}\left(y_t^4\right) = \frac{3\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)}$$

when $0 < \alpha < \frac{1}{\sqrt{3}}$. Chen and Hong (2016), Dahlhaus et al. (2006), and Inoue et al. (2020) have documented that the parameters θ are often unstable over time when this model is applied to financial data.

We accommodate time-varying parameters in this model by letting $\theta_t = (\omega_t, \alpha_t)$ and setting

$$\mathbb{E}[g(v_t, \theta_t)] = 0 \qquad \text{with} \qquad g(v_t, \theta_t) = \begin{bmatrix} y_t^2(y_t^2 - \lambda_t) \\ y_t^2 - \frac{\omega_t}{1-\alpha_t} \\ y_t^4 - \frac{3\omega_t^2(1+\alpha_t)}{(1-\alpha_t)(1-3\alpha_t^2)} \end{bmatrix} .$$

where $v_t = (y_t^2, y_t^4)$. Compared to the MA model above, this model falls more precisely in our extended GMM framework as it is nonlinear, time-varying, and over-identified.

1.3.3 Consumption Capital Asset Pricing Model

In this model, an individual can invest in J risky assets, each with returns $R_{j,t+1}$ (where j ranges from 1 to J), as well as a risk-free asset with a fixed return

$R_{f,t+1}$. The Euler equation asset pricing model assumes that an individual has a preference for consumption C that follows a power utility function.

$$U(C_t) = \frac{C_t^{1-\alpha}}{1-\alpha} ,$$

where α is a risk aversion parameter. Together the Euler equations are given by the $J+1$ nonlinear equations

$$\mathbb{E}\Big[(1+R_{f,t+1})\beta(\frac{C_{t+1}}{C_t})^{-\alpha}|I_t\Big] - 1 = 0 ,$$
$$\mathbb{E}\Big[(R_{j,t+1}-R_{f,t+1})\beta(\frac{C_{t+1}}{C_t})^{-\alpha}|I_t\Big] = 0, \quad j = 1, ..., J ,$$

where I_t represents the information available at time t and β denotes the time discount parameter. The parameters of interest are given by $\theta = (\beta, \alpha)$.

There is ample evidence that the parameters of this model might change over time. For example, Bruno and Shin (2015), Rey (2015), Guiso et al. (2018), and Miranda-Agrippino and Rey (2020) all found evidence that risk aversion changes over time synchronizing with the business cycle. The time-varying parameter GMM framework I propose in this paper allows θ to change smoothly over time and can therefore capture changes in risk aversion and possibly discounting.

1.3.4 Gravity Models

Another example of models the proposed method can handle is the gravity model for international trade, see Eaton and Kortum (2002), Anderson and Van Wincoop (2003), Silva and Tenreyro (2006). The gravity model is an analytical framework for studying bilateral trade flows. A simple version of this model is:

$$\mathbb{E}[Y_t|X_t] = \exp(X_t'\theta)$$

where Y_t denotes trade values such as nominal export, imports, or total trade between country t and the reference country, X_t contains country t features such as nominal Gross Domestic Product (GDP) and distance between the most populated city of country t and the reference country.

As pointed out by Nishihat and Otsu (2020), the estimation of this model can be viewed as a conditional GMM estimation. By the law of iterated expectations, the conditional moment restriction above implies unconditional model

restrictions of the form:

$$\mathbb{E}[\{Y_t - \exp(X_t'\theta)\}h(X_t)] = 0$$

for any function $h(.)$. In particular, $h(X) = X$ yields the Poisson pseudo maximum likelihood (PPML) estimator while $h(X) = exp(X'\beta)X$ yields the nonlinear least squared (NLS) estimator. Notice that for either choice of the function $h(X)$, the moment condition is a nonlinear function of the parameter θ.

According to Marimoutou et al. (2010) shows that there is some evidence that coefficients in gravity models can vary. Allowing the coefficient θ to change as θ_t may be advantageous. In this scenario, the moment condition becomes:

$$g(\theta_t) = (Y_t - exp(X_t'\theta_t))h(X_t)$$

In this context, it is possible to order the observations based on increasing spatial distance from the reference country as in Marimoutou et al. (2010). This ranking allows us to examine the impact of various factors, particularly spatial distance, on the variable Y_t. Using this ranking and considering the moment conditions, the model can fit into the framework described in this paper. This means that the assumption in our framework of a smooth time-varying coefficient implies that the economic phenomena represented by the θ_t vector do not vary significantly between a country and the slightly more distant country. In other words, the effect of X_t on Y_t is not significantly different between geographically close countries.

In conclusion, moving average, ARCH, consumption capital asset pricing models, and gravity models are all-time series (or could be viewed as) models used to forecast and analyze economic and financial data. When estimating the parameters of these models, the moment conditions are often nonlinear in the model parameters, and it is often useful to allow the parameters to vary over time. GMM can be applied to estimate the parameters in these models by using sample counterparts of the moment conditions. The usefulness of allowing the parameters to vary over time has been highlighted in various applications. The next sections will develop an extension of the classical GMM to consider these features.

1.4 Time-varying parameter GMM

I consider models defined by the following moment conditions

$$\mathbb{E}\left[g(v_{n,t},\theta_{n,t})\right]=0\,,\qquad \theta_{n,t}=\theta(t/n)\,,\quad t=1,\ldots,n\,,\quad n\geq 1\,, \tag{1.1}$$

where g is a $q\times 1$ known function that takes as inputs the observed random variables $v_{n,t}$ and the $p\times 1$ parameter vector $\theta_{n,t}$. The parameters may change over time according to the smooth function $\theta(\cdot):[0,1]\to\mathbb{R}^p$. My interest is in estimating $\theta(u)$ for some $u\in(0,1)$ given the triangular array $v_{n,t}$ for $t=1,\ldots,n$ with $n\geq 1$.

Following the literature on the non-parametric estimation of time-varying parameters (e.g. Robinson, 1989; Fan and Gijbels, 1996) I will generally assume that $\theta(\cdot)$ is a smooth function that can be approximated at u by a polynomial in t/n for $u\approx t/n$. Specifically, I consider the approximation

$$\begin{array}{rcl}\theta_u^*(t/n) & := & \beta_1+\beta_2(t/n-u)+...+\beta_{m+1}(t/n-u)^m/m! \\ & = & D(t/n-u)\beta\end{array}\,, \tag{1.2}$$

where $\beta=(\beta_1',\beta_2',\ldots,\beta_{m+1}')'\in\mathbb{R}^{p(m+1)}$, $D(u)=(1,u,u^2/2,\ldots,u^m/m!)\otimes I_p$. Note that $\beta_{i+1}=\theta^{(i)}(u)=\partial^i\theta(u)/\partial^i u$ for $i=0,1,...,m$. For ease of notation, I omit the dependence on u from the coefficients β. The polynomial order, i.e., m, is assumed to be fixed throughout the paper.

To account for the approximation error $\theta(t/n)-\theta_u^*(t/n)$ I use kernel weighted sample moments to formulate an estimator for $\theta(u)$ (e.g. Su et al., 2013). I consider

$$\tilde{g}_n(\beta|u)=\frac{1}{n}\sum_{t=1}^{n}K_b(t/n-u)\begin{bmatrix} g(v_{n,t},D(t/n-u)\beta) \\ g(v_{n,t},D(t/n-u)\beta)(\frac{t/n-u}{b})^1 \\ \vdots \\ g(v_{n,t},D(t/n-u)\beta)(\frac{t/n-u}{b})^m\end{bmatrix}\,, \tag{1.3}$$

with $K_b(\cdot)=K(\cdot/b)/b$, where $K(\cdot)$ is a kernel function with bandwidth parameter $b=b_n$. Assumptions on the bandwidth parameter are spelled out in Section 1.4.1 and Section 1.4.2 is devoted to the practical implementation.

Based on the kernel weight sample moments, I define the GMM estimator for β by

$$\hat{\beta}=\arg\min_{\beta\in\mathcal{B}}\tilde{Q}_n(\beta|u)\,,\qquad \tilde{Q}_n(\beta|u)=\tilde{g}_n(\beta|u)'\Omega_n(u)\tilde{g}_n(\beta|u)\,, \tag{1.4}$$

where $\Omega_n(u)$ is a positive definite $p(m+1) \times p(m+1)$ dimensional weighting matrix which may depend on u. Below I derive the optimal weighting matrix and clarify the parameter space $\mathcal{B} \subseteq \mathbb{R}^{(m+1)p}$. The resulting polynomial estimator of $\theta(u)$ is $\hat{\theta}(u) = \hat{\beta}_1$ and $\hat{\theta}^{(i)}(u) = \hat{\beta}_{i+1}$. Note that these estimators depend on the choice of the weighting matrix, which I suppress in the notation.

To facilitate the theoretical analysis of the proposed estimator, I follow Kristensen and Lee (2019) and introduce the following re-scaled version: $\hat{\alpha} = U_n\hat{\beta}$ where $U_n = \text{diag}(1, b, ..., b^m) \otimes I_p$. It is easy to verify that the re-scaled estimator $\hat{\alpha}$ satisfies

$$\hat{\alpha} = \arg\min_{\alpha \in \mathcal{A}} Q_n(\alpha|u) \,, \qquad Q_n(\alpha|u) = g_n(\alpha|u)'\Omega_n(u)g_n(\alpha|u) \,, \tag{1.5}$$

with

$$g_n(\alpha|u) = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u) \begin{bmatrix} g(v_{n,t}, D_b(t/n-u)\alpha) \\ g(v_{n,t}, D_b(t/n-u)\alpha)(\frac{t/n-u}{b})^1 \\ \vdots \\ g(v_{n,t}, D_b(t/n-u)\alpha)(\frac{t/n-u}{b})^m \end{bmatrix} \tag{1.6}$$

where $D_b(\cdot) = D(\cdot/b)$. Importantly, with this transformation, $D_b(t/n-u)$ and $K_b(\frac{t}{T}-u)$ depend on the same argument, which facilitates the derivation of precise restrictions on the parameter space $\mathcal{A}$ so that $Q_n(\alpha|u)$ is well-defined for all $\alpha \in \mathcal{A}$.

In practice, $\hat{\beta}$ can be easily computed using numerical minimization routines. However, it is useful to recall that for linear moment conditions (e.g., for a linear IV model), there exists an easy closed-form solution for problem (1.4). Further, the re-scaled $\hat{\alpha}$ is only introduced for theoretical purposes and plays no role in the practical implementation, which is discussed in section 1.4.2.

1.4.1 Asymptotic theory

I study the asymptotic properties of the class of estimators $\hat{\alpha}$ as defined in (1.5). I work in the infill asymptotic framework with rescaled time $t/n \in [0,1]$, where n denotes the number of observations. The theory is based on the assumption that the moment functions $g(v_{n,t}, \theta)$ are uniformly locally stationary (e.g. Dahlhaus et al., 1997, 2006, 2019; Kristensen and Lee, 2019). The original local stationary assumption imposes that for fixed $u \in [0,1]$, $g(v_{n,t}, \theta)$ should locally (i.e., for

$|u - n/t| << 1$) behave like a stationary process. I will effectively require the same, but uniformly over the parameter space.

To set this up, I rely on the following definition, taken from Kristensen and Lee (2019).

Definition 1. A triangular family of random sequences $W_{n,t}(\theta)$, $\theta \in \Theta$, for $t = 1, \ldots, n$ and $n \geq 1$, is uniformly locally stationary on Θ (here after $ULS(p, q, \Theta)$) for some $p, q > 0$ if there exists a family processes $W_t^*(\theta|u)$, $u \in [0, 1]$, such that: (i) The process $\{W_t^*(\theta|u)\}$ is stationary and ergodic for all $(\theta, u) \in \Theta \times [0, 1]$; (ii) for some $C < \infty$ and $\rho < 1$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \|W_{n,t}(\theta) - W_t^*(\theta|u)\|^p\right]^{1/p} \leq C\left(\left|\frac{t}{n} - u\right|^q + \frac{1}{n^q}\right) .$$

This definition establishes that the nonstationary process $W_{n,t}(\theta)$ is well approximated by its stationary approximation $W_t^*(\theta|u)$ around u.

Definition 2. A stationary process $W_t^*(\theta|u)$ is said to be L_p-continuous w.r.t θ if the following holds for all $\theta \in \Theta$: $\mathbb{E}[||W_t^*(\theta|u)||^p] < \infty$ and

$$\forall \epsilon > 0 \; \exists \delta > 0 : \mathbb{E}\left[\sup_{\theta' : ||\theta - \theta'|| < \delta} ||W_t^*(\theta'|u) - W_t^*(\theta|u)||^p\right]^{1/p} < \epsilon$$

As it is stated in Kristensen and Lee (2019), the definition of L_p-countinuous w.r.t θ is weaker than almost sure continuity, and it can be shown that it also implies stochastic equicontinuity of $Q_n(\alpha|u)$, see proof of theorem 1 in the appendix.

Definition 3. Let $\epsilon_t, t \in \mathbb{Z}$ be a sequence of i.i.d random variables. Let $\mathcal{F}_t := (\epsilon_t, \epsilon_{t-1}, ...)$ and $\mathcal{F}_t^{\star\star(t-k)} := (\epsilon_t, ..., \epsilon_{t-k+1}, \epsilon_{t-k}^{\star\star}, \epsilon_{t-k-1}, \epsilon_{t-k-2}...)$, where $\epsilon_{t-k}^{\star\star}$ is a random variable which has the same distribution as ϵ_0 and is independent of all $\epsilon_t, t \in \mathbb{Z}$. For a process $W_t = H_t(\mathcal{F}_t) \in L^q$ with deterministic $H_t : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ let define $W_t^{\star\star(t-k)} := H_t(\mathcal{F}_t^{\star\star(t-k)})$ and the uniform functional dependence measure

$$\delta_q^W(k) := \sup_{t\in\mathbb{Z}} \left\| W_t - W_t^{\star\star(t-k)} \right\|_q$$

With these definitions in place, I state the main assumptions that are needed to show point-wise consistency for the proposed nonparametric GMM estimator.

Assumption 1. *For the moment model defined by equation* (1.1)*, I assume that*

1. *(i) The kernel function* $K(.) \geq 0$ *is bounded with bounded variation* B_k *and compact support* $\mathcal{K} = [-1, 1]$ *and* $\int K(v)dv = 1$*; (ii)* $K(.)$ *is symmetric around 0; (iii) for some* $C < \infty$*,* $|K(v) - K(v')| \leq C|v - v'|$*,* $v, v' \in \mathbb{R}$*.*

2. $\mathcal{A} = \{\alpha \in \mathbb{R}^{p(m+1)} : D(v)\alpha \in \Theta, \forall v \in \mathcal{K}\}$*, where* Θ *is compact and the true value* $\theta(u) \in \Theta$*.*

3. $\Omega_n(u)$ *is a positive definite and* $\Omega_n(u) \xrightarrow{p} \Omega(u)$ *with* $\Omega(u)$ *positive definite.*

4. *(i)* $g(v_{n,t}, \theta)$ *is* $ULS(p, q, \Theta)$ *for* $p \geq 1$ *and* $q > 0$ *with stationary approximation* $g_t^*(\theta|u)$*; (ii)* $\theta \mapsto g_t^*(\theta|u)$ *is* L_1 *continuous; (iii)* $\theta \mapsto \mathbb{E}[g_t^*(\theta|u)] = 0$ *if and only if* $\theta = \theta(u)$*.*

Assumption 1.1 excludes Gaussian and other high-order kernels but includes the Epanechnikov and triangular kernels. The compact support in assumption 1.1 is related to standard assumption 2: we need Θ to be compact as is standard in the literature to obtain the desired uniform convergence results. But if the kernel has unbounded support then $D_b(v)\alpha \notin \Theta$ as $b \to 0$ for $\alpha = (\alpha_1, ..., \alpha_m)$ with $\alpha_i \neq 0$ for some $i > 1$ and any $v \neq 0$. We would need the parameter space $\mathbb{A}$ to collapse to $\{(\alpha_1, 0, ..., 0) : \alpha_1 \in \Theta\}$ in order to allow for unbounded kernel support, which would make unfeasible the Taylor approximation with respect to α of the loss function. However, by restricting the support of the kernel, we can verify that $K_b(v)f(v_t, D_b(v)\alpha)$ is well-defined for all $\alpha \in \mathbb{A}$ and $v \in \mathbb{R}$. Additionally, as we will notice in the proof, we can argue that $(\alpha_1, .., 0)$ is an interior point of $\mathbb{A}$ and then apply standard Taylor approximation in our analysis of $\hat{\alpha}$. Assumption 1.3 regarding the weighting matrix is standard in the GMM literature. Finally, Assumption 4. (i) and 4. (ii) are standards in the analysis of "global" extremum estimators of stationary models. In particular, assumption 4. iii with $K \geq 0$ guarantees that the local polynomial estimator identifies $\theta(u)$.

Theorem 1. *Given assumption 1 as* $b \to 0$ *and* $nb \to \infty$ *we have that*

$$\hat{\alpha} \xrightarrow{p} (\theta(u), 0, ..., 0)'$$

In particular, $\hat{\theta}(u) \xrightarrow{p} \theta(u)$*.*

The theorem above only provides consistency of $\hat{\theta}(u)$ but is silent about $\hat{\alpha}_{i+1}$ for $i = 1, ..., m$. However, I can provide a precise analysis of these estimates with additional regularity conditions on the moment condition function.

Assumption 2. *For the moment model defined by equation* (1.1)*, we assume that*

1. *(i)* $g(v_{n,t}, \theta_{n,t})$ *is twice continuously differentiable with respect to* $\theta_{n,t}$*; and (ii)* $\theta(u)$ *lies in the interior of* Θ *and is* $m+1$ *times continuously differentiable with respect to* u*.*

2. *For each* $u \in [0,1]$ *and* $j = 1, ..., q$*, there exists a measurable function* $H(u, .)$ *such that* $g^*_{jt}(\theta|u) = H(u, \mathcal{F}_t)$ *and* $\delta_q^{g^*_j}(k) := \sup_{u\in[0,1]} \delta_q^{g^*_j(\theta|u)}(k)$ *satisfies that* $\triangle_{0,q}^{g^*_j} := \sum_{k=0}^{\infty} \delta_q^{g^*_j}(k) < \infty$ *where* $\delta_q^{g^*_j(\theta|u)}(k)$ *is the uniform functional dependence measure. Let us define* $\Sigma(u) = \sum_{k\in\mathbb{Z}} Cov(g^*_0(\theta(u)|u), g^*_k(\theta(u)|u))$.

3. *Let be* $s(v_{n,t}, \theta) = \frac{\partial g(v_{n,t},\theta)}{\partial \theta'} \in \mathbb{R}^{q\times p}$*, then* $s(v_{n,t}, \theta)$ *is* ULS $(p, q, \{\theta : ||\theta - \theta(u)|| < \epsilon)\}$ *for some* $p \geq 1$ *and* $q, \epsilon > 0$ *with* L_1 *continuous stationary approximation* $s^*_t(\theta|u)$ *and* $\Sigma_s(u) = \mathbb{E}[s^*_t(\theta|u)]$.

Assumption 2.1 is standard in the polynomial fitting estimation literature and is needed to apply the Taylor approximation. Assumption 2.2 refers to a measure of dependence frequently used in the local stationary literature, see Dahlhaus et al. (2019). Finally, assumption 2.3 is needed to obtain an approximation for the bias and convergence of the derivatives of the moment conditions. Given these assumptions, I obtain the following result.

Theorem 2. *Suppose the assumptions 1 and 2 hold. Then, as* $b \to 0$ *and* $nb \to \infty$*,*

$$\sqrt{nb}U_n\Big\{\hat{\beta}(u) - \beta_0(u) - U_n^*[G_0(u)'\Omega(u)G_0(u)]^{-1}G_0(u)'\Omega(u)\Big(\mathbb{B}(u) + o_p(1)\Big)\Big\}$$
$$\to N\Big(0, [G_0(u)'\Omega(u)G_0(u)]^{-1}G_0(u)'\Omega(u)\Lambda(u)G_0(u)[G_0(u)'\Omega(u)G_0(u)]^{-1}\Big)$$

with $U_n^* = diag\{b^{m+1}, b^m, ..., b\} \otimes \mathbb{I}_p$*,* $\mathbb{B}(u) = \mathbb{M} \otimes \Sigma_s(u)\frac{\theta^{(m+1)}(u)}{(m+1)!}$*,* $\Lambda(u) = \mathbb{K} \otimes \Sigma(u)$*, and* $G_0(u) = \Big(\mathbb{I}_{m+1} \otimes \Sigma_s(u)\Big) \times \mathbb{J}$*, where* $\mathbb{I}_p$ *is the identity matrix of order* p*, and* $\mathbb{M}$*,* $\mathbb{K}$*, and* $\mathbb{J}$ *are matrix of constants defined as:* $\mathbb{M} =$

$diag\{\mathbb{K}_0^1, \mathbb{K}_1^1, \ldots, \mathbb{K}_m^1\} \in \mathbb{R}^{(m+1)\times(m+1)}$, $\mathbb{J} = \begin{bmatrix}\mu_0' & \mu_1' & \ldots & \mu_m'\end{bmatrix}' \in \mathbb{R}^{p(m+1)\times p(m+1)}$, *and*

$$\mathbb{K} = \begin{bmatrix} \mathbb{K}_0^2 & \mathbb{K}_1^2 & \ldots & \mathbb{K}_m^2 \\ \mathbb{K}_1^2 & \mathbb{K}_2^2 & \ldots & \mathbb{K}_{m+1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{K}_m^2 & \mathbb{K}_{m+1}^2 & \ldots & \mathbb{K}_{2m}^2 \end{bmatrix} \in \mathbb{R}^{(m+1)\times(m+1)}$$

with $\mathbb{K}_i^1 = \int K(v)v^{m+1+i}dv$, $\mathbb{K}_i^2 = \int K^2(v)v^i dv$, *and* $\mu_i = \int K(v)v^i D(v)dv$.

The asymptotic covariance can be estimated using plug-in methods. It follows from the proof of Theorem 2 and by applying propositions 2.5 and 3.8 in Dahlhaus et al. (2019) for covariances functionals using the invariance property of the stationary approximation that a consistent estimator for the asymptotic variance-covariance matrix is $\hat{\Lambda}(u) = \mathbb{K} \otimes \hat{\Sigma}(u)_{HAC}$ where:

$$\hat{\Sigma}(u)_{HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{n-1} w_{i,n}(\hat{\Gamma}_i + \hat{\Gamma}_i')$$

where

$$\hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} K_b(t/n-u)g(v_{n,t}, D_b(t/n-u)\hat{\alpha}(u))g(v_{n,t-j}, D_b((t-j)/n-u)\hat{\alpha}(u))'$$

with $\hat{\alpha}(u) \to \alpha^*(u) = (\theta(u)', \mathbf{0}', ..., \mathbf{0}')'$. Similar estimator for the covariance has been proposed in Dahlhaus (2012).

As in the standard GMM setup, we can obtain the optimal $\Omega_n(u)$ to minimize the asymptotic variance-covariance matrix. If we consider $\Omega_n(u) = \hat{\Lambda}^{-1}(u)$ where $\hat{\Lambda}(u) = \mathbb{K} \otimes \hat{\Sigma}(u)$, then the asymptotic variance-covariance matrix of $\hat{\beta}(u)$ is $[G_0(u)'\Lambda^{-1}(u)G_0(u)]^{-1}$.

1.4.2 Practical implementation

In this section, I put the pieces from the previous sections together and describe the algorithm for computing the efficient GMM estimate for $\theta(u)$. Specifically, for a given u and bandwidth b consider the following steps.

1. Pick a suboptimal weighting matrix, such as the identity matrix $\Omega_n(u) = I_{q(m+1)}$, and solve $\hat{\beta}^0 = \arg\min_{\beta\in\mathcal{B}} Q_n(\beta|u)$ for the objective function $Q_n(\beta|u) = g_n(\beta|u)'\Omega_n(u)g_n(\beta|u)$.

2. Estimate $\hat{\Lambda}(u) = \mathbb{K} \otimes \hat{\Sigma}(u)_{HAC}$ using $\hat{\beta}^0$ to estimate $\hat{\Sigma}(u)_{HAC}$ as explained in Section 1.4.1.

3. Set the optimal weighting matrix $\Omega_n(u) = \hat{\Lambda}^{-1}(u)$ and solve $\hat{\beta}^1 = \arg\min_{\beta \in \mathcal{B}} Q_n(\beta|u)$ where $Q_n(\beta|u) = g_n(\beta|u)' \hat{\Lambda}^{-1}(u) g_n(\beta|u)$.

4. The variance matrix for $\hat{\beta}^1$ is computed as $[\hat{G}_0(u)' \hat{\Lambda}^{-1}(u) \hat{G}_0(u)]^{-1}$ where we use $\hat{\beta}^1$ to estimate $\hat{G}_0(u)$.

1.4.3 Bandwidth Selection

It is well-established that the choice for bandwidth b plays a crucial role in balancing the bias and variance of kernel estimators. For time-varying parameter linear regression models, there are several approaches for estimating an optimal bandwidth parameter, i.e., mean squared error minimizing. One approach is the use of a nonparametric version of AIC (e.g. Cai and Tiwari, 2000; Cai, 2002, 2007), which has been shown to perform well compared to other methods such as generalized cross-validation and the classical AIC. Another approach is the use of cross-validation, as proposed in the work of Robinson (1989) and Chen (2015). The latter study also proposes a plug-in method for selecting the optimal bandwidth, but this method requires the estimation of the derivative of the parameter of interest.

For time-varying autoregressive models, time-varying moving average models, and time-varying GARCH models, Richter and Dahlhaus (2019) propose an adaptive bandwidth selection method based on cross-validation for local M-estimators. It should be noted that this method does not apply to the GMM estimator, which is an extremum estimator that is not an M-estimator. Therefore, cross-validation is not guaranteed to result in the optimal bandwidth parameter for locally stationary processes. To the best of my knowledge, there is no cross-validation procedure for the GMM framework. As a result, we rely on rule-of-thumb methods for selecting the bandwidth in both the simulation and empirical sections of our analysis. There is no doubt that further research is necessary for this area.

1.5 Simulation Study

In this section, I present the results of various Monte Carlo simulations designed for a MA(1) model and an ARCH(1) model. For all simulations, we set $m = 1$ and estimate the local linear estimator of the time-varying coefficient.

The mean absolute deviation error (MADE) and root mean squared error (RMSE) are considered to assess the finite-sample performance of the proposed estimator. In all cases, I utilize the Epanechnikov kernel and apply the rule of thumb guidelines for the bandwidth parameter. The MADE and RMSE are calculated as follows:

$$MADE(u) = S^{-1}\sum_{s=1}^{S}|\hat{\theta}^s(u) - \theta(u)| \quad RMSE(u) = \sqrt{S^{-1}\sum_{t=1}^{S}(\hat{\theta}^s(u) - \theta(u))^2}$$

where $\hat{\theta}^s(u)$ is the local linear estimator of $\theta(u)$ at replication s with $s = 1, 2, ..., S$. I also compute the coverage ratio of the 95% confidence interval for the relevant parameters.

1.5.1 Simulation Design

To evaluate the performance of the proposed estimator in various contexts, we consider three experiments based on different shapes for the time-varying parameters: TV_1, in which the coefficients follow a cosine shape; TV_2, in which the coefficients exhibit a linear trend with a break at $u = 0.50$; and TV_3, in which the coefficients follow a square root function.

The data generating process and moments conditions are set as follows:

1. In simulation exercise 1, I consider a time-varying MA(1) model:

   $$y_t = e_t + \theta_t e_{t-1}$$

   with $\epsilon_t \sim N(0, \sigma_t^2)$. The moments conditions considered are: $E(y_t^2) = \sigma^2(1 + \theta^2)$ and $E(y_t y_{t-1}) = \sigma^2\theta$.

2. In simulation exercise 2, I consider a simple restricted time-varying ARCH(1):

   $$y_t = \sqrt{\lambda_t}\epsilon_t\ , \quad \lambda_t = 1 - \alpha_t + \alpha_t y_{t-1}^2$$

where $\epsilon_t \sim N(0,1)$. The moments conditions I use here are $E(y_t^2(y_t^2 - \lambda_t)) = 0$ and the fourth moment $E(y_t^4 - \frac{3w^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)}) = 0$. This is particularly challenging because sometimes the sample replication for the fourth moment is not informative.

1.5.2 Results

I run 1000 Monte Carlo replications for sample sizes T=500 and T=1000 and estimate the time-varying parameters at three points in time of the sample: $u = 0.10, 0.50, 0.90$ for each of the models and experiment considered.

Tables A.1 to Table A.5 present the main results of the simulation exercises. Table A.1 and Table A.2 display the MADE and RMSE, respectively, for the tv-MA(1) model, where it can be observed that these measures improve as the sample size is increased in all three experiments. When there is an increase or decrease in performance as u increases, this is due to the different values of the true parameter at each specific value of u. Table A.3 shows that the coverage ratios fall within the acceptable range.

The restricted tv-ARCH(1) model was also analyzed, with sample sizes of T=2000 and T=4000 included to align with the typical sample sizes used in ARCH/GARCH models and to account for the estimation of a fourth moment. Table A.4 presents the performance of the nonparametric GMM estimator in terms of the MADE and RMSE for this model. It can be seen that the best performance is generally achieved at $u = 0.50$, as this is the point at which the most observations are used to estimate the parameter relative to the extreme points of $u = 0.10$ and $u = 0.90$. The coverage in Table A.5 improves quickly with increasing sample size and is generally better for the $u = 0.5$ estimation.

1.6 Empirical Study

I apply the proposed procedure in 3.3 to study a simple gravity model for international trade of the US economy, explained in section 1.3.4. In this sense, I explore "distance"- varying-parameter gravity model as in Marimoutou et al. (2010) by ordering the data according to increasing spatial distance: the first observation corresponds to the nearest neighbor of the US (Canada) and the last

one to the country very far from US (Indonesia) such that the top 100 trade partners are included. The data is obtained from the Gravity database from Research and Expertise on the World Economy CEPII for 2017. I focus on the estimation of the "distance" varying coefficients in the model by considering the following gravity model moment condition:

$$\mathbb{E}[\{Y_t - \exp(X_t'\beta_t)\}X_t] = 0$$

where $t = 1, 2, ..., 100$, Y_t is the exports of the US to the country t and X_t are the regressors for country t which includes a constant term, the logarithm of GDP and the logarithm of distance: $X_t = (1, log(GDP_t), log(distance_t))'$. The ranking enables us to study the effect of spatial distance on export relationships. This model specification fits within our general framework. This framework requires moment conditions that depend nonlinearly on the parameters, which are allowed to vary based on distance rather than time. By allowing β_t to change smoothly, it is assumed that the parameters are not significantly different between close countries geographically. We use $h(X) = X$ in the moment condition specified above as in the Poisson pseudo maximum likelihood (PPML) approach.

1.6.1 Results

After applying the nonparametric GMM estimator to study a gravity model with distance-varying coefficients for the US and main partners' datasets, this section discusses the main findings.

Figure A.1 (a) and Figure A.1 (b) display, respectively, the "distance" varying coefficients for GDP and distance in the gravity model considered for the US economy and top 100 trade partners.[4]. The first result supports previous findings in the gravity model literature: the coefficient for GDP is positive, while the coefficient for distance variables is mostly negative. The reasoning behind these signs is that the partner's GDP represents the market's attractiveness; in other words, the larger the GDP of the partner, the higher the US exports to that partner. In the case of distance, it is traditionally viewed as a proxy for the cost of entering a market. Therefore, the greater the distance, the higher the cost of entry, discouraging exports to countries far from the US. These relationships align with the expectations of the law of gravity for trade.

[4]Epanechnikov kernel and rule of thumbs for the bandwidth were used.

The plots reveal new findings in gravity models for the United States. Figure A.1(a) shows the distance-varying importance of the size of the destination market for US exports. I find evidence that the importance of the partner's GDP diminishes as the distance increases, with a coefficient of around 1.2 for countries close to the US and a coefficient of 0.5 for countries further from the US. For countries halfway between these two options, the coefficient of the GDP is around 1. It is worth noting that the coefficient of GDP (in logs) on US exports (in logs) can be interpreted as an elasticity, meaning that for countries near the US, a 1% increase in the partner's GDP leads to a 1.2% increase in US exports to that partner.

To further analyze this trend, I will divide the countries into three regions based on their distance from the US: those that are 3500 km or less away, those that are between 10000 km or less but more than 3500 km away, and those that are 10000 km or further away. In the closest region, up to 3500 km away, the coefficient on GDP (in logs) decreases in a similar trend to that of countries far away from the US, farther than 10000 km. In contrast, the countries in the second region have a constant coefficient value.

On the other hand, Figure A.1(b) also reveals new information: the coefficient for the distance variable is initially negative and decreasing for countries close to the US, then rapidly increases towards zero and stabilizes around this value. To further analyze this trend, let us divide the plot into three areas: the first one for countries up to 5000 km away from the US, the second one for countries between 7500 km or less but more than 5000 km away from the US, and the third area for countries that are 7500 km or further away from the US. Similar to the case of GDP, we can interpret this coefficient as an elasticity: for countries around 5000 km away from the US, a 1% increase in distance is associated with a 6% decrease in US exports.

The plot shows that in the first area, the distance variable becomes strongly negative, which highlights the increasing trade costs in the neighborhood of the US. In the second area, the rapid increase of the coefficient of the distance variable suggests that this variable loses relevance in explaining bilateral trade. In the last area, the coefficient fluctuates around a zero value, indicating that this variable no longer helps to explain US exports.

1.7 Conclusion

In this paper, I propose a new nonparametric GMM estimator in the presence of time-varying parameters. In the spirit of Fan and Gijbels (1996), Dahlhaus et al. (1997), and Kristensen and Lee (2019), I estimate the true time-varying parameters by polynomial approximation and show formally that the proposed estimator retains the properties of consistency and asymptotic normality of the standard GMM under the assumption of uniform locally stationarity. This concept allows us to approximate the non-stationary process locally by its stationary approximation, which facilities the application of modified versions of the law of large numbers and the central limiting theorem. In the Monte Carlo study, the proposed estimator shows good performance for the time-varying MA model and time-varying ARCH model in terms of MADE and RMSE. Coverage ratios also show to be reasonably good. In the application, I study a simple gravity model for international trade with varying parameters for the US economy and main trade partners and find evidence of varying effects of partners' GDP and distance on US exports. I let unexplored an optimal criteria to select the bandwidth parameter, which might be the topic of future research.

Chapter 2

SIMPLE FUNCTIONAL LOCAL PROJECTIONS

Abstract

I propose a simple and practical method for estimating impulse responses. The method is based on using a single flexible function to approximate the impulse response and comes with honest confidence intervals that take into account the bias that is introduced by the functional approximation. The method facilitates an easy way to interpret and compare impulse responses across model specifications and estimation methods. I demonstrate the performance of the proposed method in a simulation study and apply it to study the effects of total factor productivity shocks on the US economy. The results show that the functional local projection approach provides accurate and informative estimates for impulse response functions.

2.1 Introduction

The estimation of impulse response functions is a central task of macroeconomic and macroeconometric research. A large portion of the econometric efforts has been directed toward finding ”efficient” estimation methods, that is, studying the conditions under which certain estimators should be preferred over others. From

the macroeconomic literature, the emphasis has been on discovering "new" evidence for macroeconomic relationships, naturally using the available econometric estimation methods.

As such, there has been little effort in synthesizing and unifying the different impulse response estimates obtained. For instance, what is the peak of the effect of unemployment in response to a one-unit exogenous change in the short-term interest rate, and what is the half-life of such peak. When searching for such answers, the literature provides little guidance as impulse responses are plotted but rarely numerically quantified in terms of their properties. This makes comparisons complicated, and little research has looked into the differences across methods and methodologies.[1]

In this paper, I provide some useful tools for characterizing impulse response functions using their canonical characteristics: peak, location, and the half-lives before and after the peak. To do so, I introduce simple functional local projections as a way to summarize the evidence from any impulse response estimator. For clarity of exposition, I focus on the local projection estimates, but the methodology is applicable to any impulse response estimate.

The main idea is to use one or a few, Gaussian basis functions to quantify the key features of a given impulse response. Gaussian basis functions are attractive for this purpose as their parameters directly capture the characteristics of the impulse response and no summary statistics of the function need to be computed.

Typically, using only a single or a few basis functions to approximate the impulse response leads to biased estimates. To keep inference honest, I show how size-correct confidence bands can be constructed by taking into account both the bias and the variance of the functional approximation. These bands are conservative, but in realistic simulations, their length does not exceed the length of standard LP confidence intervals by much. The advantage is that the functional LPs are smooth and easily interpretable.

I extend the static functional local projection framework to allow for time-varying impulse responses. While there exists a large literature for time-varying impulse responses in the context of vector autoregressive models (e.g. Cogley and Sargent, 2005; Primiceri, 2005), much less work has been done for local projection models (Ruisi, 2019; Inoue et al., 2022). I develop a non-parametric

[1] A notable exception is Coibion (2012) who investigates the differences in the effects of monetary policy stemming from standard VARs, and the Romer and Romer (2004a) distributed lag approach.

estimator that allows for smooth time-variation in the impulse responses (e.g. Robinson, 1989). The estimator is shown to be consistent and asymptotically normal. Crucially, this extension enables us to study the time variation in the characteristics of the impulse responses.

In a simulation study, the performance of functional local projection and local projection estimates are compared in terms of their coverage ratios and lengths for different levels of smoothness and sample sizes. The results indicate that functional local projections perform best in terms of coverage when the degree of smoothness is high and worst when the degree of smoothness is low. Additionally, the functional local projections show improved performance in terms of coverage as the sample size increases. These results suggest that functional local projections can provide accurate and reliable estimates for certain cases. However, their performance may vary depending on the specific characteristics of the data generating process. Put differently, it is easy to construct a DGP where functional local projections are uninformative, but for commonly observed shapes of impulse responses, the Gaussian basis functions can effectively summarize the information in impulse responses.

The empirical study shows that the functional local projection approach to estimating impulse response functions yields smoother and more informative results than the traditional local projection method. The functional approach provides additional information about the impulse response functions through the estimated parameters of the Gaussian basis function, allowing for a more detailed understanding of the system dynamics. In the case of total factor productivity shocks, the functional approach indicates that their impact on GDP is significant but temporary at intermediate horizons, while the standard method suggests that their impact persists in the long term. The functional approach also reveals that the impulse response functions may be time-varying, with differences in peaks and timing of peaks at different points in the sample. Overall, the functional local projection method provides valuable practical knowledge into the effects of total factor productivity shocks on the economy. For example, the cumulative impulse response function of GDP to a total factor productivity shock is characterized by a peak value of $a = 1.1$, time to peak of $b = 14.2$, and half-lives before and after the peak of $c\sqrt{2} = 138\sqrt{2}$. If time-varying coefficients are allowed, the peak of the GDP's response to a total factor productivity shock changes over time, starting at 1.7 at the first quantile of the sample, dropping to 1.4 at the middle point, and then rising again to 1.8 at the third quantile.

The remainder of this paper is organized as follows. Section 2 describes the papers related to this study. Section 3 introduces the main functional local projection approach and clarifies the benefits of using Gaussian basis functions to get a smooth version of impulse response functions. Section 4 develops two estimation procedures to estimate the parameters for the functional local projection. Section 5 offers a procedure to obtain confidence bands considering the bias induced by the Gaussian approximation. Section 6 displays the simulation study, while Section 7 shows an extension of the functional local projection by allowing the response coefficients to change over time. Finally, Section 8 shows the empirical application. The appendix shows the proofs and assumptions for the time-varying extension and displays tables and plots for the simulation and empirical study.

2.2 Literature Review

The present work is related to estimating impulse response functions using local projection methods. The modern literature on local projections was started by Jordà (2005), see also Stock and Watson (2018). My discussion of the literature focuses on two key features of impulse response functions that can be incorporated into the local projection framework: smoothness and time instability.

Several previous studies have focused on obtaining smooth estimates of impulse response functions. Prominent examples include Plagborg-Møller (2017), Barnichon and Matthes (2018) and Barnichon and Brownlees (2019). These studies are motivated by the observation that standard local projection methods often suffer from high variability due to their nonparametric nature. As a result, the mean squared error of such estimates can be disproportionally large, implying low efficiency.

Plagborg-Møller (2017) proposes methods for optimally selecting the degree of smoothing for a general shrinkage estimator of the impulse response function and provides procedures for obtaining point-wise and joint confidence sets in this context. Barnichon and Matthes (2018) propose a functional approximation of the impulse response function using a few basis functions, which serves as a dimension reduction tool and makes the estimation more feasible. Barnichon and Brownlees (2019) introduces a B-spline smoothing technique called smooth local projections for impulse response function estimation, which increases precision while preserving the flexibility of standard local projection methods. Re-

cently, Miranda-Agrippino and Ricco (2021) proposes a Bayesian local projection method that regularizes impulse response functions using appropriately chosen priors.

There is less research on incorporating time-varying coefficients into the local projection framework. Two relevant studies in this area are Ruisi (2019) and Inoue et al. (2022). Ruisi (2019) models local projections in a time-varying framework under a Bayesian setting. Inoue et al. (2022) develops a local projection estimator for estimating impulse responses in the presence of time variation, allowing for slope coefficients and variance changes.

In this paper, we aim to contribute to the existing literature on impulse response function estimation by presenting a framework that allows for the easy incorporation of smoothness and time variation within the local projection framework.

2.3 Functional Local Projections

In this section, I introduce local projections in a general GMM framework and discuss the functional local projection approximation.

2.3.1 Local projections

A generic local projection model is given by

$$\tilde{y}_{t+h} = \tilde{x}_t \beta_h + \tilde{w}_t' \gamma_h + \tilde{\epsilon}_{t+h} \,, \qquad h = 0, 1, \ldots, H \,,$$

where $\tilde{y}_{t+h}$ is the h periods ahead outcome variable, $\tilde{x}_t$ is the explanatory variable of interest, $\tilde{w}_t$ is a vector of pre-determined control variables and $\tilde{\epsilon}_{t+h}$ is the error term. The coefficients of interest are β_h, for $h = 0, 1, \ldots, H$, where H is the largest horizon considered. The vector γ_h capture the effects of the control variables $\tilde{w}_t$.

Without loss of generality I first project out the control variables to obtain

$$y_{t+h} = x_t \beta_h + \epsilon_{t+h} \,, \qquad h = 0, 1, \ldots, H \,, \tag{2.1}$$

where $y_{t+h} = \tilde{y}_{t+h} - \mathrm{Proj}(\tilde{y}_{t+h}|\tilde{w}_t)$ and $x_t = \tilde{x}_t - \mathrm{Proj}(\tilde{x}_t|\tilde{w}_t)$.

Under suitable identifying assumptions, I can interpret $\{\beta_h\}_{h=0}^{H}$ as the structural impulse response function, which captures the effect of a one-unit change

in x_t on y_{t+h}. Many different identifying assumptions can be considered, e.g., short run, long run, sign, and external instruments; see **?** for a concise discussion. For concreteness, I concentrate on the case where the researcher has available an $L \times 1$ vector z_t of instruments that are correlated with x_t but uncorrelated with ϵ_{t+h}. Note that a special case occurs when $z_t = x_t$, which covers the case of short-run restrictions when $\tilde{w}_t$ is suitably chosen.

In the case where z_t's are truly external instruments, I refer to the model as the local projection instrumental variable (LP-IV) model. Recent examples of this setup include Jordà and Taylor (2016) and Jordà et al. (2020). In these applications, noisy proxies for structural shocks are used as instruments to estimate the effects of structural shocks on outcome variables. For example, in our empirical work, below z_t is a proxy for a total factor productivity shock (e.g. Fernald, 2012) and x_t is a measure of economic activity or labor productivity. Other examples are related to monetary policy shocks, fiscal policy shocks, oil price shocks, etc.

2.3.2 Functional approximation

In this section, I introduce a functional approximation for the impulse response $\{\beta_h\}_{h=0}^{H}$. The approximation aims to smooth out the otherwise noisy local projection estimates and enhance the interpretability of the properties of the impulse response function. Importantly, it is not of interest to obtain MSE optimal approximations; there works of Plagborg-Møller (2017), and Barnichon and Brownlees (2019) provide tools for this.

To this extent, in this paper, I parameterize the local projection coefficients with Gaussian basis functions following Barnichon and Matthes (2018). Besides smoothing out the original estimated impulse responses, this approach also benefits from the fact that the basis functions have an easy interpretation, as can be seen from the baseline expression:

$$\beta_h^f(\theta) = a \exp(-\frac{(h-b)^2}{c}) \ ,$$

where the parameters $\theta = (a, b, c)'$ can be adjusted to approximate the shape of the true impulse response function. These parameters provide important practical information about the characteristics of the impulse response function. The parameter a represents the peak or height of the function, b represents the timing

of the peak, and c captures the persistence of the shock's effect. Specifically, c is the amount of time required for the effect of the shock to be 50% of the peak value. This allows us to easily interpret the features of the impulse response function.

One possible extension of the basis function

$$\beta_h^f(\theta) = a \exp(-\frac{(h-b)^2}{c_1} 1_{h<b} - \frac{(h-b)^2}{c_2} 1_{h>b})$$

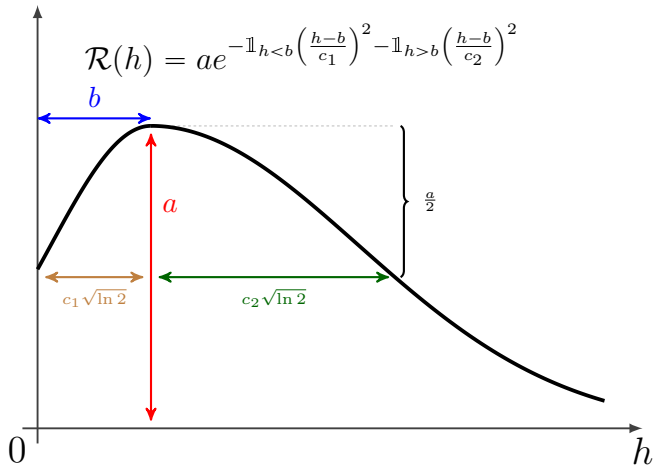
where in this case the unknown coefficients are $\theta = (a, b, c_1, c_2)$. This extension of the basis function adds elements to characterize the impulse response function compared to the original basis function. The extended Gaussian basis functions are slightly modified compared to the standard case where $c_1 = c_2$. This extension is important as it relaxes the symmetry assumption that is otherwise imposed by the Gaussian basis function. To illustrate this extension, Figure 2.1 shows a hump-shape impulse response corresponding to a particular Gaussian basis function. The coefficient a captures the magnitude of the peak of a shock, b captures the time to the peak, $c_1\sqrt{\ln 2}$ captures the "half-life" to the peak, and $c_2\sqrt{\ln 2}$ the "half-life" from the peak.

It is generally possible to use multiple basis functions rather than just one. For example,

$$\beta_h^f(\theta) = \sum_{j=1}^{N} a_j \exp(-\frac{(h-b_j)^2}{c_j}) \ .$$

When more than one basis function is used, the interpretability of the estimated parameters may no longer be guaranteed. However, in some cases, the approximation with multiple Gaussian-basis functions can still be interpretable. For example, consider an oscillating pattern, the approximation with two Gaussian functions may still be interpretable if the first Gaussian basis function captures the initial, positive hump-shaped response and the second Gaussian function captures the larger, negative hump-shaped effect that occurs later. This allows the different basis functions to capture different aspects of the response to the shock.

Figure 2.1: Gaussian Basis Function



Notes: This plot represents a typical shape of the Gaussian basis function in its extended version.

In general, I define the functional local projection model as

$$y_{t+h} = x_t \beta_h^f(\theta) + e_{t+h} \,, \qquad h = 0, 1, \ldots, H \,, \tag{2.2}$$

where the true impulse response is approximated by $\beta_h^f(\theta)$. Here e_{t+h} includes the original error term as well as the error stemming from the approximation of β_h.

2.4 Estimation

In this section, I discuss two asymptotically equivalent procedures for estimating the impulse responses based using the functional local projection model (2.2). The first one uses a generalized method of moments formulation for estimating θ, whereas the second one is based on a direct fit of the Gaussian basis function to an unrestricted local projection estimate.

2.4.1 Generalized Method of Moments

Given the parametrization of the impulse response functions explained above, I estimate the unknown coefficients θ in (2.2) using the generalized method of moments (GMM). The moment conditions are given by

$$E(z_t(y_{t+h} - x_t\beta_h^f(\theta))) = 0 \ , \quad \text{for all} \quad h = 0, \ldots, H \ .$$

Given a sample of observations for time periods $t = 1, 2, ..., n + H$, I replace the population moments with their sample counterparts. I obtain

$$g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_t(\theta) \qquad \text{where} \qquad g_t(\theta) = \begin{bmatrix} z_t(y_t - x_t\beta_0^f(\theta)) \\ z_t(y_{t+1} - x_t\beta_1^f(\theta)) \\ ... \\ z_t(y_{t+H} - x_t\beta_H^f(\theta)) \end{bmatrix} \ .$$

With this notation established, the generalized method of moments estimator for θ is given by

$$\hat{\theta} = \arg\min_{\tilde{\theta}\in\Theta} n g_n(\tilde{\theta})' \hat{W}_n g_n(\tilde{\theta})$$

where $\hat{W}_n$ is an $L(H+1) \times L(H+1)$ positive semi-definite weighting matrix. Standard efficient choices can be considered (e.g. Hall, 2005).

Given the estimates $\hat{\theta}$ I replace the unknown coefficients of the Gaussian basis functions by their estimates and recover an estimate for β_h^f. That is

$$\hat{\beta}_h^f = \beta_h^f(\hat{\theta}) = \hat{a}\exp(-\frac{(h-\hat{b})^2}{\hat{c}}) \ .$$

This functional impulse response estimate is smooth by construction, and its characteristics are easily obtained from $\hat{\theta}$.

2.4.2 Indirect fitting

An alternative approach for estimating θ is to fit the Gaussian basis function approximation directly to the unrestricted local projection estimates $\hat{\beta}$. The latter estimates can be obtained using standard IV methods based on the moment conditions $E(z_t(y_{t+h} - x_t\beta_h)) = 0$. I use this estimate to obtain an estimation for θ by solving the minimum distance problem.

$$\hat{\theta} = \arg\min_{\theta\in\Theta}(\hat{\beta} - \beta_h^f(\theta))' W (\hat{\beta} - \beta_h^f(\theta)) \ ,$$

where W is a weighting matrix for the horizons. It follows from standard results in Newey and McFadden (1994a) that this estimator is equivalent to the direct GMM estimator discussed above.

This indirect estimation procedure can be related to the framework proposed in Plagborg-Møller (2017). Specifically, he considers estimators of the form

$$\hat{\beta}(\lambda) = \arg\min_{\beta \in \mathbb{R}^n} (\beta - \hat{\beta})' W (\beta - \hat{\beta}) + \lambda (\beta - \hat{\beta}_h^f)' W (\beta - \hat{\beta}_h^f) \ ,$$

where $\hat{\beta}_h^f = \beta_h^f(\hat{\theta})$, $\hat{\theta}$ is defined above, and λ a parameter that controls the amount of shrinkage towards the Gaussian basis function.

Two important differences distinguish the approach discussed from Plagborg-Møller (2017). First, I take the shrinkage parameter to infinity, $\lambda \to \infty$, which means that I shrink the LP estimates completely toward the basis functions that I use to smooth the impulse response function. This improves the interpretation of the results due to the Gaussian basis function and saves computation time because I do not need to implement a procedure to compute an optimal shrinkage parameter. Second, In contrast to Plagborg-Møller (2017) I consider basis functions that are nonlinear in their primitive parameters. This means that the functional local projection estimator is not available in a closed-form solution, and I solve the minimum distance problem numerically.

2.5 Honest confidence bands

In this section, I describe a procedure for computing confidence bands for the functional local projection estimator $\hat{\beta}_h^f$. Crucially, the approach considers the bias introduced by approximating the impulse responses with Gaussian basis functions.

Specifically, I do not assume that $\hat{\beta}_h^f \xrightarrow{p} \beta_h^\dagger$ which would follow if $\hat{\theta} \xrightarrow{p} \theta$ and the true $\beta' s$ satisfy $\beta_h^\dagger = \beta_h^f(\theta)$. Clearly, this argument relies on the assumption that the functional approximation is exact in population, which seems too strong for practical applications. Therefore, I construct confidence bands while taking into account the bias that is introduced by the functional approximation. For this purpose, note that the mean squared error of the estimator can be decomposed as

$$\text{MSE}(\hat{\beta}_h^f) = \mathbb{E}(\hat{\beta}_h^f - \beta^\dagger)^2 = Var(\hat{\beta}_h^f) + [\mathbb{E}(\hat{\beta}_h^f) - \beta^\dagger]^2$$

where $\beta^{\dagger}$ is the true parameter and $\hat{\beta}_h^f$ is the functional local projection estimate.

This mean squared error can be easily approximated by simulation, taking advantage of the fact that the true $\beta^{\dagger}$ is consistently estimable using the unrestricted local projection estimator. In particular, suppose that:

$$\sqrt{n}(\hat{\beta} - \beta^{\dagger}) \xrightarrow{d} N(0, \Sigma)$$

where $\hat{\beta}$ is the unrestricted local projection and Σ is the variance-covariance matrix from the local projection estimator. In addition, suppose there is an estimator $\hat{\Sigma}$ such that $\hat{\Sigma} \xrightarrow{p} \Sigma$.

For a given h, confidence bands are based on computing an estimator for the MSE. First, define for any $\eta \in \mathbb{R}^{H+1}$ and $(H+1)\times(H+1)$ symmetric positive definite weight matrix W I define

$$\hat{\theta}(\eta) = \arg\min_{\theta}(\eta - \beta_h^f(\theta))' W (\eta - \beta_h^f(\theta))$$

where W is a weighting matrix for the horizons.

Construct the draws $\hat{\beta}^{(s)} \sim N(\hat{\beta}, \hat{\Sigma})$ from $s = 1, 2..., M$, and compute an estimate for the MSE error using

$$\hat{\text{MSE}}(\hat{\beta}^f) = \frac{1}{M}\sum_{s=1}^{M}\left(\beta^f(\hat{\theta}(\hat{\beta}^{(s)})) - \bar{\beta}^f\right)\left(\beta^f(\hat{\theta}(\hat{\beta}^{(s)})) - \bar{\beta}^f\right)' + (\bar{\beta}^f - \hat{\beta})(\bar{\beta}^f - \hat{\beta})'$$

where $\bar{\beta}^f = \frac{1}{M}\sum_{s=1}^{M}\beta^f(\hat{\theta}(\hat{\beta}^{(s)}))$. It is straightforward to check that $\hat{\text{MSE}}$ is a consistent estimator of the MSE.

Based on the estimate of the MSE I compute the $(1-\alpha)\%$ pointwise confidence intervals for the functional local projection estimates from

$$C_{\alpha} = \hat{\beta}_h^f \pm z_{\alpha}\sqrt{\hat{\text{MSE}}(\hat{\beta}^f)_{hh}}$$

for $h = 0, ..., H$, where z_{α} is the $1-\alpha$ quantile corresponding to the standard normal distribution.

Proposition 1. *Assuming that* $\sqrt{n}(\hat{\beta} - \beta^{\dagger}) \xrightarrow{d} N(0, \Sigma)$. *In addition, consider that we have available the estimator* $\hat{\Sigma}$ *such that* $\hat{\Sigma} \xrightarrow{p} \Sigma$. *Then as* $n \to \infty$*:*

$$Pr(\beta^{\dagger} \in C_{\alpha}) \geq 1 - \alpha$$

where C_{α} *is defined above, the proof can be found in the appendix.*

Alternatively, I consider the following $(1-\alpha)\%$ joint confidence intervals for the functional local projection estimates as

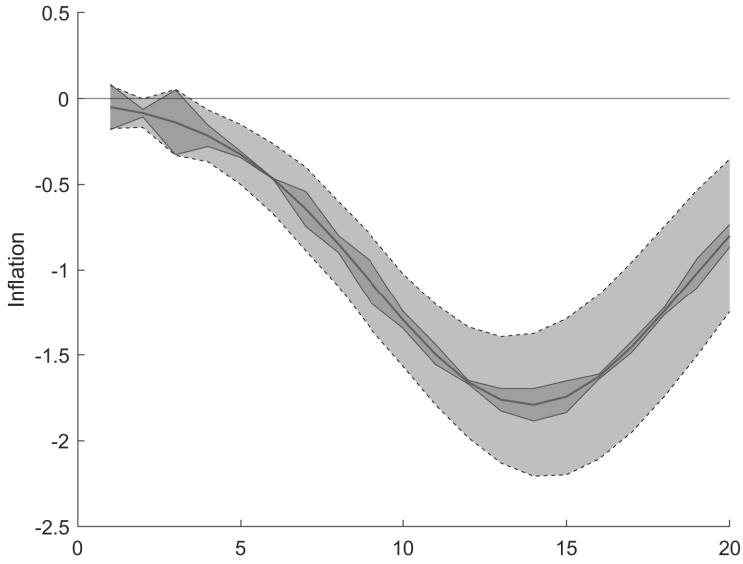
$$\hat{\beta}_h^f \pm z_\alpha \sqrt{\max_{h\in\{1,\dots,H\}} \hat{MSE}(\hat{\beta}^f)_{hh}} \,.$$

The confidence interval constructed using the maximum of the estimated MSE ensures that the resulting confidence bands have a level of at least $1-\alpha$. Additionally, this approach continues with the honest confidence bands by considering the bias introduced by the approximation in the construction of the confidence bands. Using the maximum quantity when building join confidence sets has recent applications as in **?**.

As mentioned in Plagborg-Møller (2017), in repeated experiments, pointwise confidence intervals ensure a predetermined asymptotic coverage probability for each impulse response individually. These intervals are frequently employed in applied macroeconomics and panel event studies. On the other hand, joint confidence intervals of asymptotic level $1-\alpha$ have a probability $1-\alpha$ of covering the true impulse response at all horizons in repeated experiments for large sample sizes. Joint bands are necessary for situations such as testing whether the entire impulse response function is zero or examining the overall shape of the impulse response function. Indeed, Sims and Zha (1999) and Inoue and Kilian (2016) have advocated for using joint bands rather than pointwise bands in macroeconomic applications.

Figure 2.2 shows a decomposition of a typical confidence interval constructed following the previous procedure. The black line indicates the functional local projection estimate of inflation to a monetary policy shock for the high degree of smoothness cases; see Section 2.6. The darker gray area represents the part of the confidence set due to the bias. In comparison, the lighter gray area captures the part of the confidence region due to the variance of the functional local estimator. For this particular example, the bias contribution is bigger than the variance contribution to the confidence set at the shorter horizons. This is because most of the functional projection estimates from the draws $\hat{\beta}^s$ initiate very close to zero looking to fit the "bell shape" curve around horizon 15, which causes the low variance at around horizon 1. Indeed, the variance contribution increases along the horizons.

Figure 2.2: Bias and Variance



Note: The graph displays a decomposition between bias and variance of a typical confidence interval for an impulse response function following the procedure developed in section 2.5. The darker gray area represents the part of the confidence set due to the bias, while the lighter gray area captures the part of the confidence region due to the variance of the functional local estimator.

2.6 Simulation Study

In this section, I discuss the results from a simulation study that was designed to benchmark the performance of impulse response estimation based on the functional local projection method relative to the standard local projection approach.

2.6.1 Data Generating Process

I consider the same simulation design as in Barnichon and Brownlees (2019). Specifically, I consider a standard macroeconomic system with Gross Domestic

Product (GDP) growth gdp_t, Personal Consumption Expenditure (PCE) inflation π_t, and the Fed funds rate ffr_t

$$gdp_t = \sum_{h=0}^{20} \beta_{11(h)} \epsilon_{t-h}^{gdp} + \sum_{h=1}^{20} \beta_{12(h)} \epsilon_{t-h}^{\pi} + \sum_{h=1}^{20} \beta_{13(h)} \epsilon_{t-h}^{ffr} ,$$

$$\pi_t = \sum_{h=0}^{20} \beta_{21(h)} \epsilon_{t-h}^{gdp} + \sum_{h=0}^{20} \beta_{22(h)} \epsilon_{t-h}^{\pi} + \sum_{h=1}^{20} \beta_{23(h)} \epsilon_{t-h}^{ffr} ,$$

$$ffr_t = \sum_{h=0}^{20} \beta_{31(h)} \epsilon_{t-h}^{gdp} + \sum_{h=0}^{20} \beta_{32(h)} \epsilon_{t-h}^{\pi} + \sum_{h=0}^{20} \beta_{33(h)} \epsilon_{t-h}^{ffr} ,$$

where ϵ_t^{gdp}, ϵ_t^{π}, and ϵ_t^{ffr} are i.i.d. structural normal shocks with mean zero and variances equal to σ_{gdp}^2, σ_{π}^2, and σ_{ffr}^2. Notice that the system is generated using the usual recursive order restriction such that the monetary policy shock does not affect contemporaneously neither GDP growth or inflation.

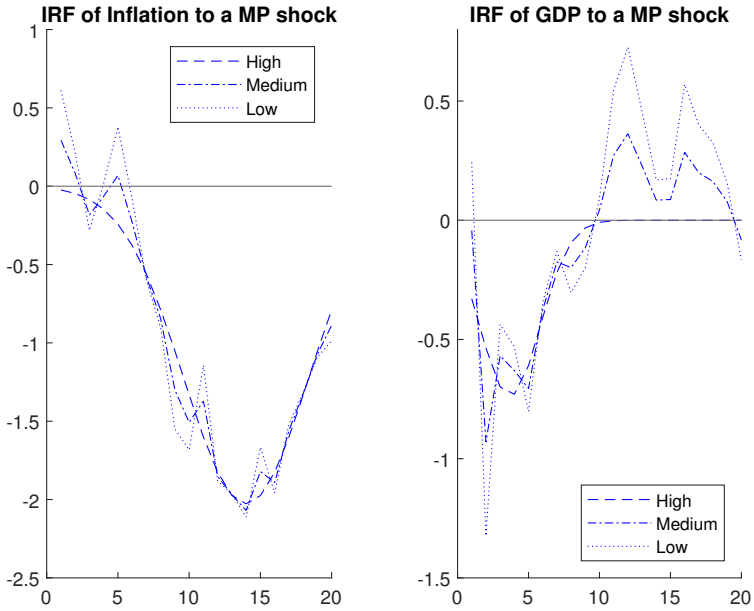
In order to have realistic dynamics in the data generating process, the impulse response parameters $\beta_{ij}(h)$ are set equal to the coefficients of the nine structural impulse response functions estimated with local projection over 1959Q1-2007Q4 for the US economy for the variables GDP growth, PCE inflation, and fed fund rates. These coefficients are estimated following Barnichon and Brownlees (2019), who identified the impulse response functions through a recursive ordering by including in the local projection regression the correct subset of control variables.

Using these estimates and the data generating process described above, we simulate the data we use to implement the functional and standard local projections methods to estimate the impulse response functions. For estimation purposes, we considered the shocks as if they were observed by the econometrician. This is a common assumption when studying the properties of impulse response estimation methods; see, for instance, Li et al. (2022) who refer to it as the observed shock identification scheme. The exercise focuses on estimating the impulse response of π_t and gdp_t to a monetary policy shock ϵ_{ffr} via functional local projections and compares this to the standard local projection estimates.

We consider two experiments. In the first experiment, we aim to evaluate the performance of the functional local projection method in relation to the degree of smoothness of the true impulse response function. To do this, we will conduct

three sets of simulations in which we manipulate the coefficients of interest (i.e., β_{13} or β_{23}) to increasingly deviate from a perfect fit to a Gaussian function to their local projection estimates. Meanwhile, the other coefficients will be set to their local projection estimates. We expect the functional local projection method to perform best when the true impulse response function is set to a Gaussian shape over the horizons.

Figure 2.3: Experiment 1 and Degree of Smoothness of the True IRF

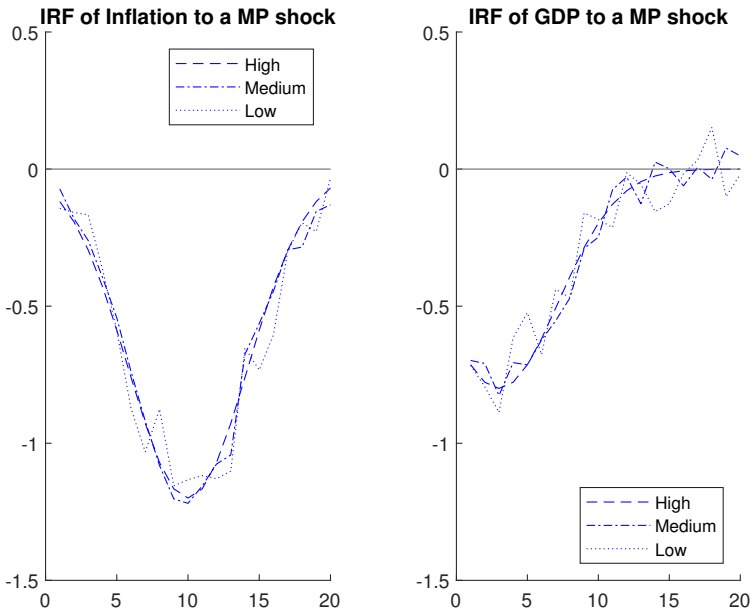


Note: The figure illustrates the true impulse response function for inflation (on the left) and GDP (on the right) in response to a monetary policy shock at various levels of smoothness in experiment 1.

In the second experiment, we will examine the performance of the functional local projection method in a more controlled environment, where the true impulse response function has a Gaussian shape with varying levels of noise. In the case of low-variance noise, the function will have a high degree of smoothness, while in the case of high-variance noise, the function will have a low degree of smoothness. This experiment aims to assess the performance of the functional

local projection method when the true impulse response function is believed to be a smooth bell-shaped function over the horizons. Therefore, I expect to observe further improvements in the functional local projection method over the traditional local projection estimation in this experiment. The true impulse response functions for each experiment are reported in Figures 2.3 and 2.4.

Figure 2.4: Experiment 2 and Degree of Smoothness of the True IRF



Note: The figure illustrates the true impulse response function for inflation (on the left) and GDP (on the right) in response to a monetary policy shock at various levels of smoothness in experiment 2.

2.6.2 Results

This section presents the results of 1000 Monte Carlo replications for sample sizes of $n = 200$ and $n = 500$, in which functional local projection and local projection estimates and their confidence sets are calculated to analyze two key features of the confidence bands: coverage ratios and lengths.

Table 2.1 and Table 2.2 show the coverage ratios and lengths of a 90% confi-

dence interval for the response of inflation and GDP growth, respectively, every horizon and specification of the degree of smoothness of the true impulse response parameters for Experiment 1 with a sample size of $n = 250$. The first half of the columns shows the coverage ratios, while the last half displays the lengths. For each of these features of the impulse response functions, the first three columns contain the results for the high degree of smoothness case, the following three represent the medium degree of smoothness case, and the last three include results for the low degree of smoothness case. For each of these cases, the first two columns display the coverage ratios or lengths for the functional local projection confidence sets denoted as FLP and FLP_j, where the subscript j represents the joint confidence interval. In addition, the next column displays the results for the standard local projection estimate, which I denominate as LP.

In general, Table 2.1 and Table 2.2 demonstrate the good performance of the confidence interval coverage ratio for functional local projection compared to standard local projection. For inflation impulse response functions, functional local projection confidence bands outperform those based on standard local projection in cases with high levels of smoothness, particularly at shorter horizons, while maintaining a minimum cost in terms of length. The low coverage around horizon 4 is due to the jump for the jagged true impulse response function into positive territory, which the functional local projection with one Gaussian function is unable to capture. These issues are not present in experiment 2, as shown in Table B.3 to Table B.6 for sample sizes of $n = 250$ and $n = 500$ available in the appendix. In these cases, functional local projection confidence bands demonstrate competitive coverage and length across smoothness and horizons. Additional results can be found in the appendix of this chapter, which includes plots of specific simulations for the functional local projection and its confidence intervals for experiment 1 with a sample size of $n = 500$.

The subsequent subsections provide a more in-depth examination of the results of the simulation design, in which the performance of functional and standard local projection methods was compared. The first subsection focuses on the relative performance of these methods based on the degree of smoothness of the true impulse response function. The second subsection compares the performance of both methods at different horizons of the impulse response functions in more detail.

Table 2.1: Coverage and Length for 90 Confidence Interval with n=250 in Experiment 1

	Inflation																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**
1	0.97	1.00	0.84	0.90	1.00	0.83	0.97	1.00	0.83	0.31	1.19	0.32	1.13	1.49	0.36	2.15	2.39	0.44
2	0.96	1.00	0.85	0.76	1.00	0.83	0.75	1.00	0.83	0.39	1.19	0.45	0.62	1.49	0.50	0.96	2.39	0.57
3	0.96	1.00	0.84	0.80	1.00	0.83	0.68	1.00	0.85	0.45	1.19	0.51	0.50	1.49	0.55	0.70	2.39	0.63
4	0.93	1.00	0.85	0.90	1.00	0.82	0.78	1.00	0.83	0.50	1.19	0.59	0.61	1.49	0.62	0.80	2.39	0.70
5	0.92	1.00	0.84	0.83	0.98	0.82	0.92	0.99	0.82	0.56	1.19	0.67	1.22	1.49	0.69	2.18	2.39	0.74
6	0.88	0.99	0.83	0.86	1.00	0.83	0.86	1.00	0.84	0.60	1.19	0.68	0.79	1.49	0.68	1.19	2.39	0.71
7	0.82	0.99	0.82	0.82	1.00	0.82	0.85	1.00	0.83	0.65	1.19	0.69	0.60	1.49	0.68	0.62	2.39	0.71
8	0.81	0.97	0.82	0.78	1.00	0.81	0.77	1.00	0.82	0.70	1.19	0.70	0.68	1.49	0.70	0.76	2.39	0.76
9	0.81	0.96	0.81	0.81	0.98	0.81	0.89	1.00	0.82	0.75	1.19	0.69	1.07	1.49	0.67	1.76	2.39	0.69
10	0.84	0.94	0.79	0.80	0.98	0.81	0.82	1.00	0.81	0.79	1.19	0.70	0.95	1.49	0.70	1.35	2.39	0.75
11	0.81	0.93	0.79	0.91	0.97	0.80	0.94	0.99	0.81	0.81	1.19	0.77	1.11	1.49	0.78	1.74	2.39	0.85
12	0.80	0.91	0.77	0.79	0.97	0.80	0.80	1.00	0.80	0.85	1.19	0.86	0.86	1.49	0.86	0.94	2.39	0.93
13	0.78	0.89	0.78	0.81	0.96	0.78	0.82	1.00	0.81	0.92	1.19	0.92	0.93	1.49	0.92	1.00	2.39	0.98
14	0.78	0.86	0.77	0.78	0.92	0.77	0.79	0.99	0.80	0.99	1.19	0.96	1.01	1.49	0.96	1.11	2.39	1.04
15	0.78	0.83	0.75	0.86	0.95	0.78	0.90	0.99	0.79	1.06	1.19	1.00	1.14	1.49	1.00	1.45	2.39	1.06
16	0.78	0.82	0.77	0.77	0.89	0.79	0.79	0.99	0.78	1.09	1.19	1.03	1.12	1.49	1.02	1.25	2.39	1.07
17	0.77	0.81	0.76	0.79	0.93	0.77	0.83	1.00	0.80	1.08	1.19	1.03	1.08	1.49	1.01	1.17	2.39	1.06
18	0.76	0.82	0.77	0.75	0.93	0.76	0.80	1.00	0.79	1.04	1.19	1.03	1.03	1.49	1.01	1.10	2.39	1.07
19	0.76	0.85	0.76	0.73	0.95	0.76	0.77	1.00	0.79	0.99	1.19	1.05	0.96	1.49	1.02	1.03	2.39	1.07
20	0.79	0.91	0.77	0.68	0.97	0.75	0.68	1.00	0.78	0.96	1.19	1.06	0.91	1.49	1.04	1.04	2.39	1.09

Note: The table presents data on the coverage ratios and lengths of 90% confidence intervals for the response of inflation in experiment 1 with a sample size of $n = 250$. The data is broken down by horizon and degree of smoothness and shows results for three levels of smoothness: high, medium, and low. The coverage ratios and lengths are displayed in separate halves of the table. They are shown for functional local projection confidence sets (FLP and FLPj) and standard local projection estimates (LP). The subscript j in FLP_j represents the joint confidence interval.

Table 2.2: Coverage and Length for 90 Confidence Interval with n=250 in Experiment 1

	GDP																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**	**FLP**	**FLP$_j$**	**LP**
1	0.90	1.00	0.83	0.96	1.00	0.85	0.82	1.00	0.83	0.28	0.72	0.25	0.92	1.54	0.25	1.45	2.97	0.30
2	0.91	1.00	0.84	0.98	1.00	0.82	0.86	0.99	0.84	0.28	0.72	0.27	1.31	1.54	0.27	2.44	2.97	0.32
3	0.90	1.00	0.82	0.90	1.00	0.84	0.83	1.00	0.82	0.32	0.72	0.29	0.58	1.54	0.32	0.95	2.97	0.40
4	0.88	0.99	0.84	0.89	1.00	0.82	0.86	1.00	0.82	0.35	0.72	0.31	0.50	1.54	0.34	1.11	2.97	0.42
5	0.85	0.99	0.83	0.87	1.00	0.81	0.91	1.00	0.84	0.36	0.72	0.33	0.58	1.54	0.37	1.48	2.97	0.46
6	0.81	0.99	0.83	0.87	1.00	0.82	0.82	1.00	0.84	0.36	0.72	0.34	0.43	1.54	0.39	0.68	2.97	0.47
7	0.87	1.00	0.81	0.91	1.00	0.82	0.87	1.00	0.82	0.34	0.72	0.35	0.40	1.54	0.36	0.49	2.97	0.42
8	0.94	1.00	0.81	0.68	1.00	0.83	0.71	1.00	0.83	0.35	0.72	0.34	0.39	1.54	0.36	0.74	2.97	0.42
9	0.97	1.00	0.82	0.74	1.00	0.84	0.70	1.00	0.86	0.34	0.72	0.34	0.33	1.54	0.36	0.59	2.97	0.40
10	0.99	1.00	0.84	0.86	1.00	0.85	0.75	1.00	0.86	0.34	0.72	0.34	0.38	1.54	0.36	0.46	2.97	0.41
11	0.99	1.00	0.81	0.85	1.00	0.83	0.85	1.00	0.83	0.34	0.72	0.35	0.99	1.54	0.37	1.65	2.97	0.43
12	1.00	1.00	0.83	0.87	1.00	0.80	0.85	1.00	0.84	0.33	0.72	0.35	1.25	1.54	0.38	2.15	2.97	0.44
13	1.00	1.00	0.83	0.77	1.00	0.84	0.75	1.00	0.83	0.33	0.72	0.35	0.79	1.54	0.39	1.29	2.97	0.47
14	1.00	1.00	0.83	0.74	1.00	0.83	0.74	1.00	0.86	0.32	0.72	0.35	0.44	1.54	0.40	0.67	2.97	0.50
15	1.00	1.00	0.84	0.74	1.00	0.84	0.69	1.00	0.84	0.32	0.72	0.35	0.43	1.54	0.39	0.66	2.97	0.49
16	1.00	1.00	0.86	0.80	1.00	0.85	0.84	1.00	0.84	0.30	0.72	0.35	0.95	1.54	0.39	1.68	2.97	0.47
17	1.00	1.00	0.87	0.71	1.00	0.86	0.75	1.00	0.84	0.31	0.72	0.36	0.66	1.54	0.40	1.16	2.97	0.48
18	1.00	1.00	0.87	0.71	1.00	0.86	0.72	1.00	0.84	0.31	0.72	0.36	0.57	1.54	0.40	0.97	2.97	0.49
19	1.00	1.00	0.87	0.77	1.00	0.86	0.70	1.00	0.86	0.31	0.72	0.36	0.40	1.54	0.40	0.56	2.97	0.50
20	1.00	1.00	0.86	0.74	1.00	0.87	0.74	1.00	0.87	0.32	0.72	0.37	0.41	1.54	0.41	0.68	2.97	0.50

Note: The table presents data on the coverage ratios and lengths of 90% confidence intervals for the response of GDP growth in experiment 1 with a sample size of $n = 250$. The data is broken down by horizon and degree of smoothness and shows results for three levels of smoothness: high, medium, and low. The coverage ratios and lengths are displayed in separate halves of the table. They are shown for functional local projection confidence sets (FLP and FLPj) and standard local projection estimates (LP). The subscript j in FLP_j represents the joint confidence interval.

Degree of Smoothness

The results from Tables 2.1 and 2.2 show that the functional local projection performs best in terms of coverage ratio when the degree of smoothness is high and worst when the degree of smoothness is low, as expected. If the true parameter is already a Gaussian function, then a Gaussian approximation should perform reasonably well, as the results indicate. However, the medium and low degree of smoothness cases for the GDP is particularly challenging for the functional local projection, as they involve both negative and positive values for the true parameters as we move forward along the horizons. This explains the undercover for these cases in the second half of the considered horizons.

By reviewing Tables B.1 and Table B.2, which present results for a sample size of $n = 500$, we notice that there is an improvement in coverage for most cases, except for GDP at medium and low levels of smoothness. This could be due to the design of experiment 1, in which the true impulse response function is challenging to approximate using only one Gaussian basis function accurately. This is also evident in the results for experiment 2 shown in Tables B.3 and B.4, which have a sample size of $n = 250$. These tables demonstrate significant improvements in coverage for both inflation and GDP across all levels of smoothness. These improvements are even more pronounced in Tables B.5 and B.6, which show results for a sample size of $n = 500$.

Regarding the lengths of the 90% confidence sets, I notice some interesting points. Firstly, the lengths for the joint confidence set for the functional local projection are always higher than those of the functional or standard local projections. This is due to the definition of the joint confidence set, which considers the maximum of the mean squared errors across all horizons to compute the confidence interval. Secondly, this increase in length is the cost of achieving high coverage, as the confidence intervals consider the bias induced by the functional approximation. In this sense, I can consider these confidence intervals to be "honest".

Short vs Long Horizons

Table 2.1 shows that in the case of inflation, the best performance for the functional local projection is at the shortest horizons and deteriorates as we move forward over the horizons. The reason behind this behavior is the shape of the true impulse response functions: it is close to zero at the shorter horizons. Then it

goes down, reaching its minimum value at around horizon 15 before going back toward the zero region. Therefore, the functional local projection estimate captures this by estimating a negative value for the a parameter of around -2, a value for the b parameter of around 15, and a value for the c parameter of around 30, as is displayed in the histograms in Figure B.7 for the high degree of smoothness case.

Table 2.2 shows the opposite face regarding this behavior. In the case of GDP growth, the best performance of the functional local projection is at the longer horizons compared to the shorter horizons. Again, this result is due to the simulation design of the true impulse response function. In this case, the true function jumps into the negative territory for a short period of horizons at the beginning and then vanishes away very fast toward zero. The functional local projection estimate captures this by computing a negative value of the a of around -0.6, a value of the b parameter of around 3, and a value of the c parameter of around 15, which is shown in Figure B.7 for the high degree of smoothness case. Therefore, in this case, the coverage ratio for horizons longer than 9 is one because the functional local projection estimates and the true parameters are zeros.

In this section, the performance of functional local projection and local projection estimates were compared in terms of their coverage ratios and lengths for different levels of smoothness, sample sizes, and experiments. The results indicate that functional local projections perform best in terms of coverage when the degree of smoothness is high and worst when the degree of smoothness is low. Additionally, the functional local projections show improved performance in terms of coverage as the sample size increases. In terms of length, the joint functional local projections confidence bands consistently are longer than the alternative, which can be attributed to their definition as a joint confidence set across all horizons. The results also show that functional local projections perform best at shorter horizons, but their performance decreases as the horizon lengthens. These results suggest that functional local projections can provide accurate and reliable estimates for certain cases, but their performance may vary depending on the specific characteristics of the data.

2.7 Time-Varying Functional Local Projections

In this section, I extend the functional local projection framework to allow for time-varying impulse responses. This extension is important as there is substantial evidence for time variation in macroeconomic impulse responses (e.g. Cogley and Sargent, 2005; Primiceri, 2005). To implement this extension, I first extend the standard local projection to allow for time-varying parameters following the local polynomial approach of Fan and Gijbels (1996).

2.7.1 Time-Varying Local Projections

Consider the baseline local projection model with time-varying parameters.

$$y_{t+h} = x_t \beta_{t,h} + \epsilon_{h,t+h} \ ,$$

where $\beta_{t,h}$ is the time t horizon h impulse response.

I propose to estimate $\beta_{t,h}$ using a local linear GMM estimator. This nonparametric approach has several advantages over a parametric approach. For example, it imposes fewer restrictions on the functional form of the coefficient. In contrast, a parametric model requires assumptions about the underlying process of the coefficients. However, these assumptions are often challenging to verify so that a non-parametric approach can be more robust. Additionally, a non-parametric estimation of time-varying parameters can provide helpful information about the shape of the coefficients for subsequent parametric estimations.

The nonparametric GMM local linear estimator is based on the methodology presented in Chapter 1. However, unlike Chapter 1, the moment conditions are linear for the local projection model. This allows us to use a polynomial fitting approximation for the time-varying parameters. I proposed using local linear approximation because, typically, it is preferred over local constant models in the nonparametric literature due to their lower bias and fewer issues with boundary points (see Cai (2007), Chen and Hong (2012), among others).

I assume that these time-varying coefficients are functions of the index t scaled by the total number of periods n, i.e., I write $\beta_{t,h} = \beta_h(t/n)$. The intuition behind this assumption is that the amount of information has to increase locally as the sample size increases in order to reduce the bias and variance asymptotically. I assume that $\beta_h(t/n)$ can be approximated by a linear function, i.e., a local linear polynomial approximation, at any fixed point in time $u \in [0; 1]$

as follows

$$\beta_h(t/n) = \alpha_{0,h} + \alpha_{1,h}(t/n - u) \ ,$$

where the dependency of $\alpha_h = (\alpha_{0,h}, \alpha_{1,h})'$ on u is omitted for simplicity. The local projection regression can be written around u as

$$y_{t+h} = \tilde{x}_t \alpha_h + \epsilon_{h,t+h} \ ,$$

where $\tilde{x}_t = [x_t \quad x_t(t/n - u)]$ and $\alpha_h = [\alpha_{0,h} \quad \alpha_{1,h}]'$. The moment conditions can be written as

$$E(\tilde{z}_t(y_{t+h} - \tilde{x}_t \alpha_h)) = 0 \ ,$$

where $\tilde{z}_t = [z_t \quad z_t(t/n - u)]'$. If we stack all the moments in one vector, we get

$$\tilde{g}_t(\alpha) = \begin{bmatrix} \tilde{z}_t(y_t - \tilde{x}_t \alpha_0) \\ \tilde{z}_t(y_{t+1} - \tilde{x}_t \alpha_1) \\ ... \\ \tilde{z}_t(y_{t+H} - \tilde{x}_t \alpha_H) \end{bmatrix} = \tilde{Z}_t(Y_{t+H} - \tilde{X}_t \alpha) \ ,$$

where $Y_{t+H} = (y_t, y_{t+1}, ..., y_{t+H})'$, $\tilde{Z}_t = I_{H+1} \otimes \tilde{z}_t$, $\tilde{X}_t = I_{H+1} \otimes \tilde{x}_t$ and I_{H+1} is the $(H+1) \times (H+1)$ identity matrix. We define the local linear local projection estimator as

$$\hat{\alpha} = \arg\min\{\Big(\frac{1}{n}\sum_{n=1}^{T} K_b(t/n - u)\tilde{g}_t(\alpha)\Big)' W \Big(\frac{1}{n}\sum_{n=1}^{T} K_b(t/n - u)\tilde{g}_t(\alpha)\Big)\}$$

where the function $K_b(t/n - u)$ is defined as $\frac{K(\frac{t/n-u}{b})}{b}$ where $K(.)$ is a kernel function with bandwidth parameter b which satisfies that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. The estimator has a closed-form solution given by

$$\hat{\alpha} = (S'_{ZX} W S_{ZX})^{-1} S'_{ZX} W S_{ZY}$$

where

$$S_{ZY} = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n - u)\tilde{Z}_t Y_{t+H}$$

$$S_{ZX} = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n - u)\tilde{Z}_t \tilde{X}'_t$$

the local linear estimate for $\beta_{t,h}$ at u is $\hat{\beta}_h(u) = \hat{\alpha}_{0,h}$ for $h = 0, ..., H$. This estimator is consistent (Theorem 1.1) and asymptotically normal (Theorem 1.2) under a general set of assumptions. Assumptions and proofs are described in the appendix section.

2.7.2 Functional approximation

This section demonstrates how to extend the functional approximation method to handle time-varying parameters, as revised in Section 2.6. I will use an indirect functional local projection approach, which involves first estimating the time-varying parameter at a specific value of u, and then fitting a Gaussian basis function to these estimates. It is worth noting that an alternative approach would be to use the nonlinear GMM method introduced in Chapter 1 to directly estimate the coefficients of the Gaussian approximation for each value of u. However, I have chosen the first approach in order to demonstrate the benefits of the time-varying local projection extension designed explicitly for linear models, as described in Section 2.6.

Given the local linear estimate $\hat{\beta}(u) = (\hat{\beta}_0(u), ..., \hat{\beta}_{H+1}(u))'$ and asymptotic covariance matrix $\hat{\Sigma}(u)$ we can define:

$$\hat{\theta}(u) = \arg\min_{\theta \in \Theta} (\hat{\beta}(u) - \beta_h^f(\theta))' W (\hat{\beta}(u) - \beta_h^f(\theta))$$

such that the local linear functional local projection estimates can be defined as $\hat{\beta}_h^f(u) = \beta_h^f(\hat{\theta}(u))$. We can implement the inference procedures explained in the previous section to obtain confidence sets for every u of interest.

For example, for a given h and u, confidence bands are based on computing an estimator for the MSE of $\hat{\beta}_h^f(u)$. First, as in section 2.5 define that for any $\eta \in \mathbb{R}^{H+1}$ and $(H+1) \times (H+1)$ symmetric positive definite weight matrix W:

$$\hat{\theta}(\eta) = \arg\min_{\theta} (\eta - \beta_h^f(\theta))' W (\eta - \beta_h^f(\theta))$$

where W is a weighting matrix for the horizons.

Construct the draws $\hat{\beta}^{(s)}(u) \sim N(\hat{\beta}(u), \hat{\Sigma}(u))$ from $s = 1, 2..., M$, and compute an estimate for the MSE error using

$$\hat{\text{MSE}}(\hat{\beta}^f(u)) = \frac{1}{M}\sum_{s=1}^{M}\left(\beta^f(\hat{\theta}(\hat{\beta}^{(s)}(u))) - \bar{\beta}^f(u)\right)\left(\beta^f(\hat{\theta}(\hat{\beta}^{(s)}(u))) - \bar{\beta}^f(u)\right)' + (\bar{\beta}^f(u) - \hat{\beta}(u))(\bar{\beta}^f(u) - \hat{\beta}(u))'$$

where $\bar{\beta}^f(u) = \frac{1}{M}\sum_{s=1}^{M}\beta^f(\hat{\theta}(\hat{\beta}^{(s)}(u)))$. By the consistency of $\hat{\beta}(u)$, which is shown in the appendix, it is straightforward to check that $\hat{\text{MSE}}$ is a consistent estimator of the MSE.

Based on the estimate of the MSE I compute the $(1-\alpha)\%$ pointwise confidence intervals for the local linear functional local projection estimates from

$$C_\alpha(u) = \hat{\beta}_h^f(u) \pm z_\alpha\sqrt{\hat{\text{MSE}}(\hat{\beta}^f(u))_{hh}}$$

for $h = 0, ..., H$, where z_α is the $1-\alpha$ quantile corresponding to the standard normal distribution.

Proposition 2. *Assuming that* $\sqrt{n}(\hat{\beta}(u) - \beta(u)) \xrightarrow{d} N(0, \Sigma(u))$. *In addition, consider we have available the estimator* $\hat{\Sigma}(u)$ *such that* $\hat{\Sigma}(u) \xrightarrow{p} \Sigma(u)$. *Then as* $n \to \infty$*:*

$$Pr(\beta(u) \in C_\alpha(u)) \geq 1 - \alpha$$

where C_α *is defined above. The proof can be found in the appendix.*

Alternatively, I consider the following $(1-\alpha)\%$ local joint confidence intervals for the local linear functional local projection estimates as

$$\hat{\beta}_h^f(u) \pm z_\alpha\sqrt{\max_{h\in\{1,...,H\}}\hat{MSE}(\hat{\beta}^f(u))_{hh}}\,.$$

These intervals are local versions of those presented in section 2.5.

2.8 Empirical Study

In this section, the functional local projection estimator is applied to study the effects of technology shocks on the US economy. Specifically, I study the effects of total factor productivity shocks on US aggregate variables such as GDP, consumption, hours worked, employment, and inflation using functional local projections.

2.8.1 Total Factor Productivity shocks

The question of whether total factor productivity shocks affect variables such as GDP, consumption, or employment has been broadly discussed in applied macroeconomic research.

Seminal papers such as Gali (1999) and Basu et al. (2006) discuss the effects of technology shocks on US aggregate variables using different methodologies and find evidence that a positive total factor productivity shock leads to a decline in hours worked and employment. These findings can be explained in modern macro models by imperfect competition and sticky prices.

Francis et al. (2004) finds that technology shocks have had different effects on hours worked in the US depending on the time period considered. A positive technology shock increases hours worked in the pre-WWII period, but the opposite is found in the post-WWII period. This difference is often attributed to changes in the effect of technology shocks on productivity, i.e., the shocks were different across periods. In the early period, productivity increased immediately following a technology shock. However, in the later period, it increased gradually, providing an incentive to reduce hours worked in the short term. Francis et al. (2014) develops a method for imposing long-run restrictions in analyzing the effects of technology shocks on hours worked and confirms his previous findings.

In contrast, Uhlig (2004) finds evidence of neutral and small positive responses of hours worked and employment after a technology shock. He explains his findings through differences in taxation and long-run shifts in the social attitude towards the workplace. Along the same line, Mertens and Ravn (2011) used a vector error correction model to study the effects of exogenous tax shocks on hours worked, controlling for taxes and allowing for cointegration. They found that a total positive factor productivity shock (technology shock) increased hours worked in the short run. Similar results have been found by Alexopoulos (2011).

In sum, the literature has yet to reach a conclusion. These differences in results may be due to various factors, including the methodologies used, the time period under study, and the specific mechanisms at play. To sort things out, I argue that the functional local projection estimator can provide valuable practical information on the topic.

There are many different measurements of total factor productivity shocks

in the literature of diverse types[2]. In this empirical application, the analysis is based on the quarterly measure for total factor productivity constructed by Fernald (2014), which is a utilization-adjusted total factor productivity series labeled as "neutral total factor productivity shock". These are shocks to the total factor productivity process in the aggregate production function.

2.8.2 Specifications

This empirical application considers three different specifications to study the effects of total factor productivity shocks on key macroeconomic variables in the US economy.

The first specification uses total factor productivity as a direct observed shock, and therefore it is used as a local projection regressor. This means that total factor productivity is treated as an exogenous variable that affects the other variables in the model, and regression analysis is used to estimate this effect.

The second specification views total factor productivity shocks as instrumental variables. In this case, the true shocks are not directly observed but only through its noisy proxy (total factor productivity shocks), which is used as an instrument to help estimate the effect of the labor productivity variable (the endogenous variable) on one of the key macroeconomic variables. This is useful when the endogenous variable is correlated with the error term, which can bias the estimates. By using total factor productivity shocks as an instrument, the estimates can be made more accurate and reliable.

The third specification contemplates the case where the impulse response function coefficients are allowed to vary over time under the instrumental variable specification. In this case, allowing for time-varying coefficients can provide a more realistic and accurate representation of the effects of total factor productivity shocks on the other variables in the model. This is an extension of the functional local projection method, which was explained in detail in the previous section.

Overall, this empirical application uses three different specifications to study the effects of total factor productivity on key macroeconomic variables in the US. The first specification uses total factor productivity as a direct observed shock, the second specification uses total factor productivity as an instrumental variable, and the third specification allows the impulse response function coefficients to

[2]For an extensive list of these measures, see Ramey (2016).

vary over time. These different specifications can potentially provide a more comprehensive and nuanced understanding of the effects of total factor productivity on the economy.

To this extent, the empirical model considered in this application for the observed shock case is represented as:

$$y_{t+s} = \beta_h x_t + \epsilon^h_{t+h}$$

for $h = 0, 1, ...20$. The y_t represents the main macroeconomic variables for the US economy. They are defined as the first difference in the logarithm of GDP, consumption, hours worked, employment, and inflation. The x_t variable represents the utilization-adjusted total factor productivity growth taken from Fernald (2014).

Often applied researchers are interested in the cumulative impulse response functions, which show the sum of all the individual responses over a specified time period. For example, if a shock to a variable has a positive effect on another variable in the short term but a negative effect in the long term, the cumulative impulse response functions would show the net effect of the shock on the second variable over the entire time period. Therefore, in order to compute the cumulative responses, the following regression is considered:

$$\sum_{s=0}^{h} y_{t+s} = \beta^c_h x_t + \epsilon^h_{t+h}$$

where the subindex c refers to "cumulative" responses.

In the specification where the total factor productivity shocks are used as an instrumental variable, the regression is expressed as:

$$y_{t+s} = \beta_h x_t + \epsilon^h_{t+h}$$

with cumulative regression displayed as:

$$\sum_{s=0}^{h} y_{t+s} = \beta^c_h x_t + \epsilon^h_{t+h}$$

In this case, the endogenous variable x_t is a proxy of labor productivity which is constructed as in Gali (1999) by dividing GDP by hours worked or GDP by employment. The productivity measure and total factor productivity

shock appear to have a strong relationship, as indicated by a correlation of 0.70. This suggests that there is no weak instrument problem or a situation in which the instrument used in the regression analysis is not strong enough to identify the relationship between the variables accurately. The author of the study asserts that the total factor productivity shocks are exogenous, meaning that it is not affected by other variables in the model. As a result, the orthogonality condition should be satisfied, allowing for the use of instrumental variable regression analysis within the standard linear frameworks.

In the last specification, the impulse response coefficients are allowed to change over time. In this case, the local projection regression is specified as:

$$y_{t+s} = \beta_{t,h} x_t + \epsilon^h_{t+h}$$

with cumulative version as:

$$\sum_{s=0}^{h} y_{t+s} = \beta^c_{t,h} x_t + \epsilon^h_{t+h}$$

For all three specifications, a Gaussian basis function is used to compute the functional local projection under the framework with the idea of getting a smooth version of the impulse response function, as is explained throughout the paper in the estimation section. In particular, the parametrization used implied that the response coefficients are specified as $\beta_h = ae^{-\frac{(h-b)^2}{c}}$ or $\beta^c_h = ae^{-\frac{(h-b)^2}{c}}$. The direct estimation approach is followed in order to get the functional local projection estimates from standard local projection coefficients. The 90% confidence intervals are computed based on the MSE obtained from 1000 draws following the simulations procedures in the inference section. For all specifications, 4 lags of y_t and x_t are included as controls variables $w_t = (y_{t-1}, ..., y_{t-4}, x_{t-1}, ..., x_{t-4})$.

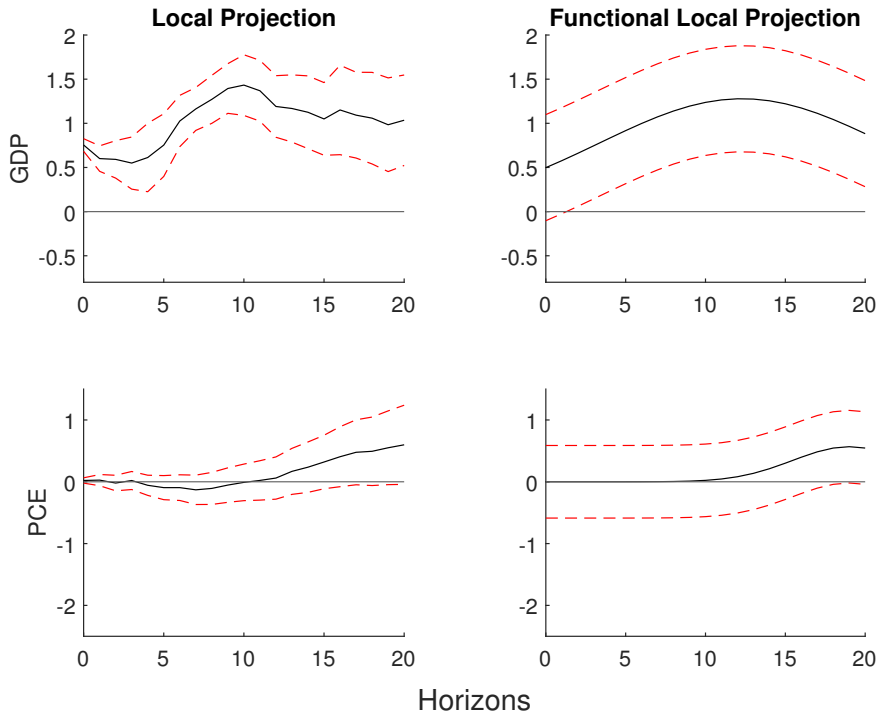
2.8.3 Results

This section presents the results from the empirical application for the three specifications and all key macroeconomic variables for the US economy for the 1947Q2-2007Q2 period, which makes the sample size $T = 241$.

Figure 2.5 and Figure B.10 present the cumulative impulse response function estimates for the local projection and functional local projection methods, respectively, for GDP, inflation, consumption, hours worked, and employment in

response to a total factor productivity shock under the observed shock specification. The first column shows the standard local projection estimates, while the second column shows the functional local projection estimates. For each impulse response function, a 90% confidence band is plotted. In the case of the standard local projection, HAC standard errors are used to calculate the confidence intervals, while for the functional local projection case, the simulation procedure described previously is used to compute the joint confidence sets. It is worth noting that the axes are set to the same values for plots in the same row to facilitate comparison.

Figure 2.5: Cumulative IRFs to a TFP shock - Observed Shock Specification - 1



Note: The figure displays the cumulative impulse response function estimates for GDP and inflation in response to a total factor productivity shock, using both local projection and functional local projection methods and under the observed shock specification. The first column represents the standard local projection estimates, while the second column shows the functional local projection estimates. A 90% confidence band is plotted for each impulse response function. Regressions include 4 lags of y_t and x_t as control variables $w_t = (y_{t-1}, ..., y_{t-4}, x_{t-1}, ..., x_{t-4})$.

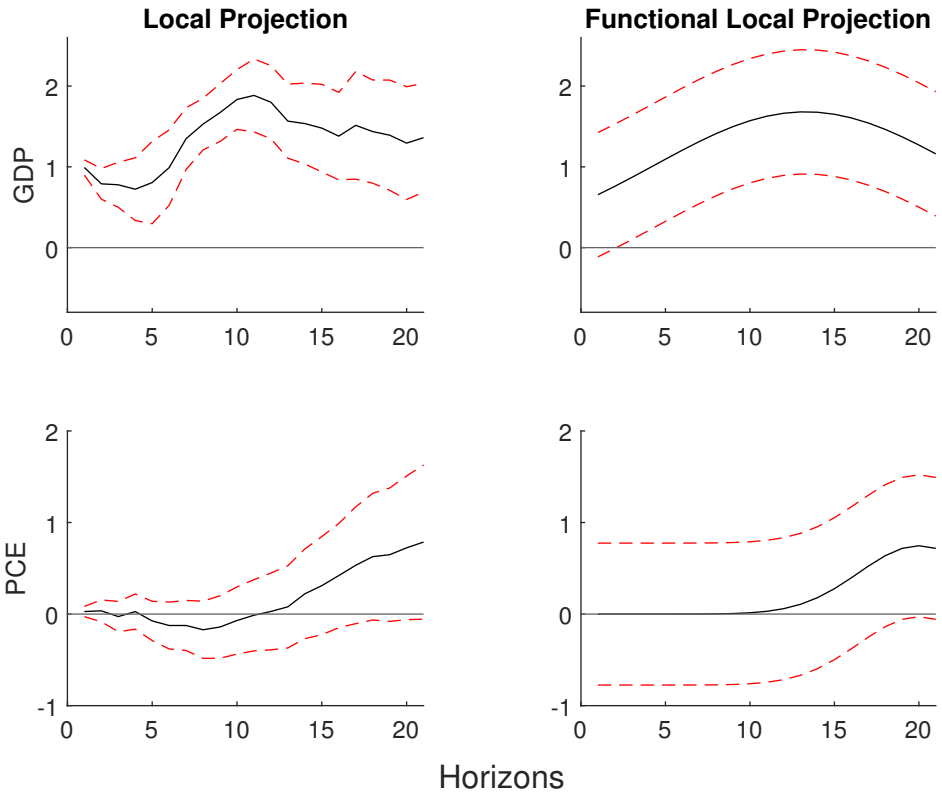
These plots demonstrate the beneficial properties of using the functional local projection approach to estimate impulse response functions. While the standard local projection estimates of the impulse response exhibit jagged, erratic behavior, particularly in the case of GDP, the functional local projection-estimated responses are smooth and follow the Gaussian parametrization. Furthermore, the functional local projection method provides additional information about the im-

pulse response functions through the estimated parameters of the Gaussian basis function, as summarized in Table B.7. In particular, the IRF for GDP is characterized by a peak value of $a = 1.1$, time to peak of $b = 14.2$, and half-lives before and after the peak of $c\sqrt{2} = 138\sqrt{2}$.

The results of the inference analysis indicate that total factor productivity shocks have no impact on consumption, hours worked, and employment in the short term. In contrast, they have a positive but transitory effect in the intermediate horizons. Both methods, LP and functional local projection agree with this. However, the standard local projection approach suggests that the initial positive effects of total factor productivity shocks on GDP persist in the long term. In contrast, the functional local projection approach finds a significant but persistent and temporary positive cumulative effect on GDP at intermediate horizons. In the case of inflation, both methods indicate that there is no impact in the short term that seems to become positive in the long horizon but remains statistically zero.

Figure 2.6 and Figure B.11 present the cumulative IRFs estimates for the local projection and functional local projection methods, respectively, in response to a total factor productivity shock under the IV shock specification. The results are similar than those obtained under the observed shock specification. In terms of the features of the impulse response function obtained from the functional local projection method, the observation is that there are small changes in the peak, while time to the peak and half-lives before and after the peak remains the same.

Figure 2.6: Cumulative IRFs to a TFP shock - IV Shock Specification - 1

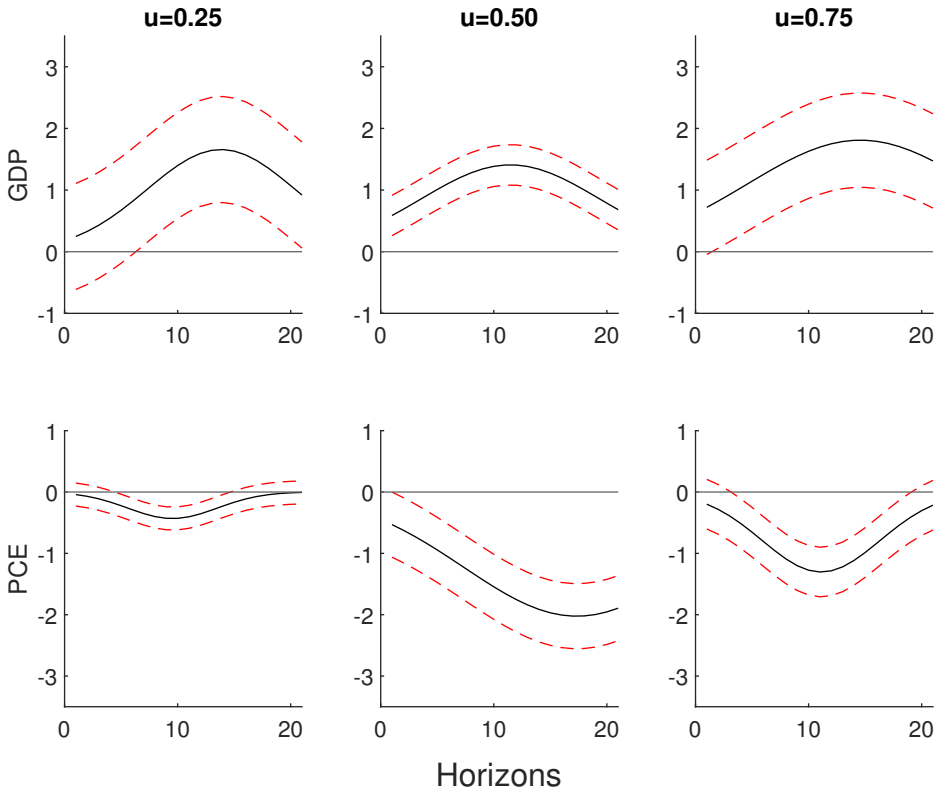


Note: The figure presents the cumulative impulse response function estimates for the local projection and functional local projection methods for GDP and inflation in response to a total factor productivity shock under the IV shock specification. The first column shows the standard local projection estimates, while the second column shows the functional local projection estimates. For each impulse response function, a 90% confidence band is plotted. Regressions include 4 lags of y_t and x_t as control variables $w_t = (y_{t-1}, ..., y_{t-4}, x_{t-1}, ..., x_{t-4})$.

Figure 2.7 and Figure B.12 show the impulse response estimates under the functional local projection method for the specification, which considers the IV with time-varying parameters. The first column shows the results for the earliest period (at the first quantile), the second column shows the results for the middle

period (at the second quantile), and the third column shows the results for the latest period (at the third quantile). Again, the axes have the same scale, so the results can be easily compared.

Figure 2.7: Cumulative IRFs to a TFP shock - IV Shock +TV Specification - 1



Note: The figure presents the cumulative impulse response function estimates for the functional local projection method for GDP and inflation in response to a total factor productivity shock under the IV+time varying coefficients shock specification. The first column shows the standard local projection estimates, while the second column shows the functional local projection estimates. For each impulse response function, a 90% confidence band is plotted. Regressions include 4 lags of y_t and x_t as control variables $w_t = (y_{t-1}, ..., y_{t-4}, x_{t-1}, ..., x_{t-4})$.

These graphs provide practical-valuable results. The impulse response functions may be time-varying because they show differences in peaks and timing of peaks at different points in the sample. For example, Table B.7 shows that the peak of the GDP's impulse response function has changed over time, starting at 1.7 at the first quantile of the sample, dropping to 1.4 at the middle point, and then rising again to 1.8 at the third quantile. This indicates that the peak of a total factor productivity shock on the GDP may vary depending on the time period.

I have included the impulse response functions for the non-cumulative cases in the appendix. These plots can be found in Figure B.13 to Figure B.18. In Table B.8, you can also find the estimates for the parameters a, b, c corresponding to this case.

2.9 Conclusion

This paper proposes a new method to smooth the impulse response functions based on functional local projection. The functional local projection brings not only smoothness to the usually jagged standard local projection estimates but also makes the result interpretable. The functional local projection uses the (Gaussian) basis function to estimate local projections while imposing smoothness on the impulse response coefficients. While different smoothing techniques are possible, i.e., P-splines, the Gaussian function has a unique feature: the coefficients of the basis function give important information regarding the impulse response function dynamics. This could be very attractive for macroeconomics because of the new knowledge they can infer with the functional local projection, such as the height of the impulse response, timing to the maximum effect, and persistence of the effect.

I present an inference procedure that considers the bias induced by the functional approximation of the impulse response and shows evidence of its good performance relative to the standard local projection inference procedure with a simulation study. In addition, for the estimation of time-varying local projection specification, functional local projection offers a flexible nonparametric alternative to parametric approaches, with the advantage of being computationally not different from kernel regression. For this extension of the functional local projection, novel results are shown on the consistency and asymptotic normality of the proposed estimator.

In the empirical section, the functional local projection approach to impulse

response function yields smoother and more informative results than the traditional local projection method. The functional local projection method provides additional information about the impulse response functions through the estimated parameters of the Gaussian basis function, allowing for a more detailed understanding of the dynamics of the system. In the case of total factor productivity shocks, the functional local projection approach indicates that their impact on GDP is significant but temporary at intermediate horizons. In contrast, the standard local projection method suggests that their impact persists in the long term. Additionally, the functional local projection method reveals that the impulse response functions may be time-varying, with differences in peaks and timing of peaks at different points in the sample. In sum, the functional local projection method provides valuable practical knowledge into the effects of total factor productivity shocks on the economy.

Chapter 3

VARMA BASED IMPULSE RESPONSE ESTIMATION

Abstract

I study the accuracy of impulse response estimates obtained from the vector autoregressive moving average (VARMA) model and compare these to the more common vector autoregressive (VAR) and local projection (LP) estimates. The evaluation of each approach is based on quantifying the bias and variance across Monte Carlo replications. The simulation design is based on a large-scale dynamic factor model known to accurately describe US macroeconomic time series. The results indicate that the VARMA models face a more significant tradeoff between bias and variance than VAR models relative to LP models, i.e., VARMA models register lower bias but higher variance at intermediate horizons.

3.1 Introduction

The estimation of impulse response functions (IRFs) has been an important topic in econometrics, and various methods have been proposed in the literature. On one side, the family of vector autoregressive methods introduced by Sims (1980) and its extension, such as vector autoregressive (VAR), structural vector autoregressive (SVAR), bayesian vector autoregressive (BVAR), structural vector au-

toregressive with IV (SVAR-IV). On another side, the family of Local Projection methods introduced by Jordà (2005) and its extensions, such as local projection (LP), local projection with IV (LP-IV), and smooth local projection (SLP). Recently, Plagborg-Møller and Wolf (2021) showed that LP and VAR methods produce similar results asymptotically, while Li et al. (2022) found that LP estimators have lower bias but higher variance compared to VAR estimators in a simulation study for thousands of relevant data generating processes (DGPs).

In recent years, there has been growing interest in using vector autoregressive moving average (VARMA) models for analyzing multivariate time series data, where theses models have been shown to be promising candidates to joint the above groups for the analysis of IRFs. Important theoretical contributions aiming at making VARMAs accessible to applied macroeconomists are due to Athanasopoulos and Vahid (2008a), Metaxoglou and Smith (2007), Poskitt (2016), Chan and Eisenstat (2017) among others. In applied work, many studies have used the VARMA method for analyzing IRFs in a specific context, such Das (2003), Raghavan et al. (2009), and Raghavan et al. (2016). Probably, the most promising area where there is more evidence of the importance of VARMA models is forecasting. In this topic, studies such as Athanasopoulos and Vahid (2008b), Dufour and Pelletier (2022), Dias and Kapetanios (2018), Athanasopoulos et al. (2012), Kascha and Trenkler (2015) have shown the good forecasting properties of VARMAs.

Despite this, VARMA models are scarcely used to represent multivariate time series compared to VAR models because the latter models are easier to implement. Indeed, VAR models can be estimated by least-squares methods, and specification is also easier for VAR models because only one lag order needs to be chosen. However, VAR models have some drawbacks. They are often less parsimonious than VARMA models, and the family of VAR models is not closed under marginalization and temporal aggregation (Fry et al. (2005)). In fact, models in macroeconomics often contain an MA component (for some examples Chen et al. (2017)), and linearized dynamic stochastic general equilibrium (DSGE) models typically result in VARMAs, not VARs (for more details, see Cooley and Dwyer (1998), Fernández-Villaverde et al. (2007), Ravenna (2007), Kapetanios et al. (2007), Chari et al. (2008)).

In consequence, VARMA models appear to be preferable from a theoretical viewpoint, but their adoption is complicated by identification and estimation difficulties. The unrestricted VARMA representation is not identified, and we need

to decide on a set of constraints to impose to achieve identification. Regarding this, Cooley and Dwyer (1998) mention: *"While VARMA models involve additional estimation and identification issues, these complications do not justify systematically ignoring these moving average components, as in the SVAR approach"*.

In this context, the aim of the present study is to conduct a simulation analysis to compare the performance of IRFs estimation from a VARMA model to the performance of the SVAR and LP estimates. The evaluation of each estimation method is based on quantifying the bias and variance across Monte Carlo replications. The simulation design is based on the work of Li et al. (2022), who randomly selected DGPs from a large-scale dynamic factor model (DFM) of Stock and Watson (2016) known to accurately describe US macroeconomic time series. For each DGP, the SVAR, VARMA, and LP methods are implemented to estimate the structural IRFs under three common identification schemes in applied macroeconomics: observed shocks, recursive, and IV/proxy identifications scheme.

I find evidence that the VARMA model exhibits a more extreme bias-variance trade-off than the VAR models. This means that at intermediate horizons, the VARMA model has a lower bias than the VAR model but higher variance at the intermediate horizon. Also, VARMA models face a similar trade-off between bias and variance to VAR models when compared with LP methods. These results suggest that the VARMA model may be a useful estimator in certain contexts where a more extreme trade-off between bias and variance is desired.

The paper is organized as follows. Section 2 presents a literature review on VARMA models and highlights the main recent contributions to specification, estimation, and identification. Section 3 shows the VARMA models and how we can obtain IRFs from the estimation of this model. Section 4 explains the simulation design. Section 5 shows the main results from the Monte Carlo exercise, and Section 6 concludes with the main takeaways and further steps.

3.2 Literature Review

This section examines previous studies on the use of VARMA models for analyzing multivariate time series data, with a focus on their ability to capture the dynamic relationships between variables in a time series and their usefulness in estimating IRFs. Previous studies have demonstrated the advantages of using

VARMA models for IRFs and forecasting compared to VAR models and have proposed new methods for identifying and estimating VARMA models. Additionally, the flexibility of VARMA models has been demonstrated in a variety of real-world applications. However, some studies have also highlighted the challenges of using VARMA models, such as their potential for non-identification and the need for careful model specification. This literature review discusses these findings and their implications for the use of VARMA models in impulse response analysis.

In recent years, there has been growing interest in using VARMA models for analyzing multivariate time series data, where these models have been shown to be promising candidates for the analysis of IRFs. Dufour et al. (2022) demonstrate the advantages of using VARMA models for impulse response estimation and forecasting compared to standard VAR models. They propose new methods for estimating and identifying VARMA models and show their effectiveness through simulation studies and an application to a six macroeconomic variable model of US monetary policy. The results show that VARMA models can improve impulse response estimation and forecasting compared to VAR models.

Some studies have addressed some limitations in the VARMA framework with researchers developing practical methods for identifying and estimating VARMA models[1]. Athanasopoulos and Vahid (2008a) proposes an extension to scalar component methodology of Tiao and Tsay (1989) for the identification and estimation of VARMA models. This leads to an exactly identified system of equations that is estimated using full information maximum likelihood. Metaxoglou and Smith (2007) introduce a state-space representation for VARMA models that enables maximum likelihood estimation using the EM algorithm. They show via simulations that the proposed algorithm converges reliably to the maximum, whereas gradient-based methods often fail because of the highly nonlinear nature of the likelihood function.

Moreover, Poskitt (2016) develops a new methodology for identifying the structure of VARMA time series models. The analysis proceeds by examining the echelon form and presents a fully automatic, strongly consistent, data-driven approach to model specification. In the context of a Bayesian approach, Chan et al. (2016), Chan and Eisenstat (2017), and Chan et al. (2022) develop diverse techniques to estimate VARMA models efficiently and demonstrate how it can

[1]For a textbook analysis of identifying and estimation method for VARMA models refer to Lütkepohl (2005).

be extended to models with time-varying vector moving average (VMA) coefficients and stochastic volatility.

Another advantage of VARMA models is their flexibility, as they can be used to model a wide range of time series data in various real-world settings. This has been demonstrated in a number of papers, including Raghavan et al. (2009), Raghavan et al. (2016), and Das (2003). The former used VARMA models to investigate Malaysian monetary policy. The authors use full information maximum likelihood to efficiently identify and estimate the model parameters and compare the impulse responses generated by VARMA, VAR, and SVAR models for the pre-and post-crisis periods. The results show that the VARMA model impulse responses are consistent with economic theories and policies pursued by the Malaysian government, particularly in the post-crisis period.

In terms of forecasting, VARMA models have been shown to be competitive with alternative methods, such as SVARs. For example, in their simulation study, as in Athanasopoulos and Vahid (2008b) and Dufour and Pelletier (2022) compare the out-of-sample forecasts generated by VARMA and VAR models and find that the VARMA models perform well. VARMA models can generate forecasts superior to those obtained from Bayesian VARs and factor models (see Dias and Kapetanios (2018). Other examples are Athanasopoulos et al. (2012), and Kascha and Trenkler (2015), while more examples can be found in Lütkepohl (2006).

Despite these theoretical developments, there are some results that point out VARMA estimation might not bring the expected improvements over VARs models. For example, Yao et al. (2017) simulated synthetic data from known data generating processes that are commonly used in economics. They compared the performance of fitted VAR and VARMA models in characterizing these data generating processes. Their results showed that while VARMA models can be theoretically identified and can produce precise estimates of impulse responses when given sufficient data, their parameters are close to the non-identified region in the parameter space. This makes it unlikely that VARMA models can produce precise estimates of impulse responses with the small amount of data typically used in macroeconomic analysis. As a result, VARMA models do not offer any significant advantage over VAR models in characterizing known data generating processes in small data sets. Another study by Kascha and Mertens (2009) estimated VARMA and state space models using simulated data from a dynamic stochastic general equilibrium model. They compared the true and estimated im-

pulse responses of hours worked in response to a technology shock. Their results showed that there were few benefits to using VARMA models.

Overall, the literature suggests that VARMA models are a potentially good candidate for the analysis of IRFs, as they are flexible, capable of capturing dynamic relationships in time series data and have been applied successfully in both theoretical and real-world settings. Additionally, recent work has addressed some of the limitations of VARMA models, making them more practical for use in empirical studies. Further research is needed to explore the potential of VARMA models for studying weakly stationary processes and developing more efficient and effective estimation methods.

3.3 VARMA Models

This section presents a simple framework for studying VARMA models with the aim of using them to estimate IRFs. Let's consider K-dimensional vector process $Y_t = (y_{1t}, ..., y_{Kt})'$ follows a VARMA(p,q) representation:

$$Y_t = \sum_{i=1}^{p} \Phi_i Y_{t-i} + U_t - \sum_{j=1}^{q} \Theta_i U_{t-j}$$

where $U_t = (u_{1t}, ..., u_{Kt})'$ are assumed to be a zero-mean sequence of uncorrelated random variables with a nonsingular $K \times K$ covariance matrix Σ_u. We can express this using matrix lag operators:

$$\Phi(L)Y_t = \Theta(L)U_t$$

where $U_t = (u_{1t}, ..., u_{Kt})'$ are assumed to be a zero-mean sequence of uncorrelated random variables with a nonsingular $K \times K$ covariance matrix Σ_u, we can express this using matrix lag operators:

$$\Pi(L)Y_t = U_t \quad ; \quad Y_t = \Psi(L)U_t$$

where $\Psi(L) = \Phi(L)^{-1}\Theta(L) = I_k - \sum_{j=1}^{\infty} \Psi_j L^j$

To gain an understanding of the identification of VARMA models, we can consider a more general representation, in which Φ_0 and Θ_0 are not identity matrices, usually refers as nonstandard VARMA. This allows us to better understand the role of these matrices in the identification of VARMA models:

$$\Phi_0 Y_t = \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} + \Theta_0 U_t - \Theta_1 U_{t-1} + ... - \Theta_q U_{t-q}$$

We can rewrite the previous equation in terms of the polynomial functions operators:

$$Y_t = (\Phi_0 - \Phi_1 L - ... - \Phi_p L^p)^- 1(\Theta_0 - \Theta_1 L - ... - \Theta_q L^q)U_t$$

Provided that Φ_0 and Θ_0 are nonsingular, this process has a standard representation. Let the previous expression be premultiply by Φ_0^{-1} and define $\bar{U}_t = \Phi_0^{-1}\Theta_0 U_t$

$$Y_t = \Phi_0^{-1}\Phi_1 Y_{t-1} + ... + \Phi_0^{-1}\Phi_p Y_{t-p} + \bar{U}_t - \Phi_0^{-1}\Theta_1\Theta_0^{-1}\Phi_0\bar{U}_{t-1} + ... \\ -\Phi_0^{-1}\Theta_q\Theta_0^{-1}\Phi_0\bar{U}_{t-q}$$

The issue of identifying unique parameterizations of VARMA models has been an important topic of study in econometrics and statistics. We say that two VARMA representations are equivalent if the MA operator $\Phi(L)^{-1}\Theta(L)$ are the same. In contrast to the reduced form VAR models, setting $\Phi_0 = \Theta_0 = I_k$ is not sufficient to ensure a unique VARMA representation. To ensure the uniqueness of a VARMA representation, we must impose restrictions on the AR and MA operators so that, for given $\Psi(L)$, there is one and only one set of operators $\Phi(L)$ and $\Theta(L)$ which generate $\Psi(L)$. This is typically done by setting a set of parameters equal to ones/zeros.

Let us consider a VARMA(1,1) model as an example of the computation of IRF. Following Tsay (2013), it is possible to achieve block identifiability for VARMA(1,1) under conditions 1: $\Phi(L)$ and $\Theta(L)$ are left-coprime and condition 2: rank of the join matrix $[\Phi_1 \quad \Theta_1] = K$, the number of variables. In this case, the IRFs can be recovered in the following way:

$$Y_t = \Phi_1 Y_{t-1} + U_t - \Theta_1 U_{t-1}$$

or $(I_K - \Phi_1 L)Y_t = (I_K - \Theta_1)U_t$. Therefore, the VMA($\infty$) representation of this process is:

$$Y_t = (I_K - \Phi_1 L)^{-1}(I_K - \Theta_1 L)U_t \\ = (I_K + \Phi_1 L + \Phi_1^2 L^2 + ...)(I_K - \Theta_1 L)U_t = \sum_{j=0}^{\infty} \Psi_j U_{t-j} \tag{3.1}$$

From the previous equation, we can recover the Ψ matrices from which the IRFs can be computed. In particular, $\Psi_0 = I_K$ and $\Psi_i = \Psi_1^{i-1}(\Psi_1 - \Theta_1)$.

3.4 Simulation

This simulation study relies on Li et al. (2022) (LPW hereafter), where the authors conduct a comprehensive simulation study, applying LP and VAR methods to thousands of empirically relevant data generating processes (DGPs). The results show that the usual least-squares LP estimator tends to have a lower bias than the least-squares VAR estimator, but this bias reduction comes at the cost of higher variance. Therefore, unless researchers are primarily concerned with bias, least-squares LP is not optimal, and shrinkage estimation via Bayesian VARs or penalized LPs is usually more attractive. This study highlights the importance of considering both bias and variance when choosing an impulse response estimator in finite samples.

In a separate paper, Plagborg-Møller and Wolf (2021) prove that LP and VAR methods asymptotically estimate the same impulse responses when the lag length used for estimation tends to infinity. However, the question of which estimator to choose in finite samples remains open.

To study the performance of estimation methods in small samples, I ran a Monte Carlo simulation study based on a small-scale version of the simulation design found on LPW in which the authors randomly selected thousands DGPs from a large-scale dynamic factor model (DFM) of Stock and Watson (2016) known to accurately describe US macroeconomic time series. For each DGP, the SVAR, VARMA, and LP methods are implemented to estimate the structural IRFs under three common identification schemes in applied macroeconomics: observed shocks, recursive, and IV/proxy identifications scheme.

Specifically, for each of the Monte Carlo replications, one DGP (instead of 3000 thousand in the original version) is randomly selected, containing three variables with a sample size of $T = 250$ from the 207-time series generated by the DFM. The estimation procedures are implemented by imposing two lags for the SVAR and LP methods, i.e., an SVAR(2) and an LP with two lags as control variables in order to mop up residual variance. In the case of the VARMA model, I consider a VARMA(2,1) specification for its parsimonious and potential to capture underlying dynamics between variables (Benati et al. (2020), Dias and Kapetanios (2018), among others). The estimation procedure for VARMA is based on maximum likelihood estimation following Tsay (2013).[2].

[2]I implement the estimation with the R package "MTS" for the VARMA models, which is a kind of companion toolbox for the textbook Tsay (2013). For the rest

The identification frameworks considered are three schemes commonly implemented in applied econometrics research as explained and applied in different works such as Romer and Romer (2004b), Ramey (2011), Gertler and Karadi (2015), Christiano et al. (1999), Blanchard and Perotti (2002)). The first is the directly observed shock scheme, in which the researcher observes both the endogenous variables and the structural shock. The second is IV/proxy identification, where the researcher uses a noisy proxy variable to infer the shock. The third scheme is one in which the researcher only observes the endogenous variables and uses recursive orthogonalization to estimate the impulse responses of the variables. These schemes are relevant in empirical studies of structural shocks, such as monetary and fiscal shocks.

3.4.1 Large-Scale Dynamic Factor Model (DFM)

The DGPs are randomly selected from an encompassing model that is known to accurately describe the population of the US macroeconomic time series: the large-scale DFM of Stock and Watson (2016). The DFM establishes that a $n_X \times 1$ vector of observed macroeconomic time series, X_t, is driven by a $n_f \times 1$ vector of latent factors, f_t, and a $n_X \times 1$ vector of idiosyncratic components v_t. The latent factors are assumed to follow a stationary VAR(p) process.

$$f_t = \Phi(L) f_{t-1} + H\epsilon_t$$

where $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{n_f t})'$ is an $n_f \times 1$ vector of aggregate shocks, which are i.i.d. and mutually uncorrelated, $Var(\epsilon_t) = I_{n_f}$. The $n_f \times n_f$ matrix H determines the contemporaneous responses of the factors with respect to the aggregate shocks. The observed macroeconomic aggregates X_t are given by

$$X_t = \Lambda f_t + v_t$$

where the idiosyncratic component $v_{i,t}$ for macro observable $X_{i,t}$ follows the AR(q) process with errors i.i.d. across t and i. All shocks and innovations are assumed to be jointly normal and homoskedastic. In order to calibrate the DFM, I follow LWP and set the number of factors $n_f = 6$ with two lags in the factor equation and two lags in the idiosyncratic component.

---

of the simulation implementation, I use the codes of LPW, which are available here https://github.com/dake-li/lp˙var˙simul

The matrix H in the factor equation varies depending on the identification scheme. LPW explains that in the case of observed shock and IV, H is chosen in a way that maximizes the effects of the first shock on fed fund rates (in the case of a monetary policy shock) and government spending (in the case of a fiscal policy shock). This ensures that monetary and fiscal shocks play a significant role in short-term changes in nominal interest rates and government spending, respectively. In the case of recursive identification, for monetary policy DGPs, the federal funds rate is placed last in the order of variables. This means that the other included variables cannot respond immediately to the monetary innovation. For fiscal policy DGPs, the government expenditure series is placed first, which means that the fiscal authority responds to other innovations in the recursive VAR with a lag.

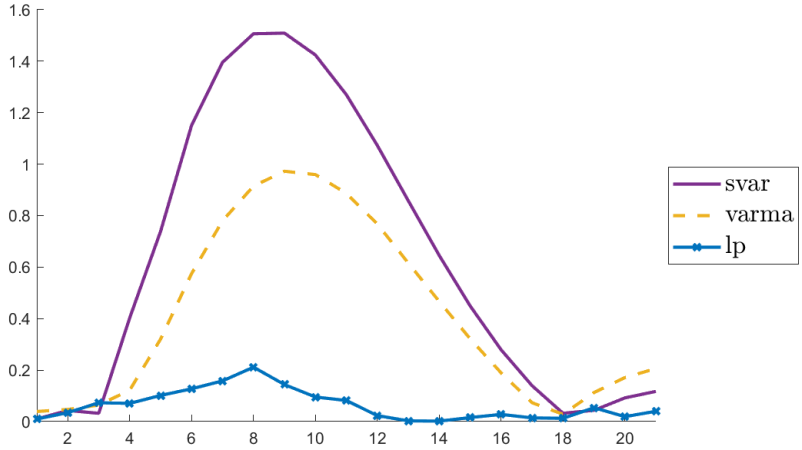
3.4.2 Small DGPs and Experiments

In this study, I investigate two sets of DGPs and examine how different estimation methods perform when applied to each of these sets under three different identification schemes. This means that a total of six experiments are conducted. The first set of DGPs includes a government spending variable and is used to study the effects of fiscal policy shocks. This set is referred to as "G". The second set of DGPs includes the fed funds rate and is used to study the effects of monetary policy shocks. This set is referred to as "MP". The remaining variables in each set are randomly selected from a pool of 207-time series variables generated using a DFM model. For each experiment, I run 1000 Monte Carlo simulations. In this small-scale simulation study, I only include one specification with three variables for each type of policy shock (as opposed to the 3000 specifications used in the original LWP paper).

3.5 Results

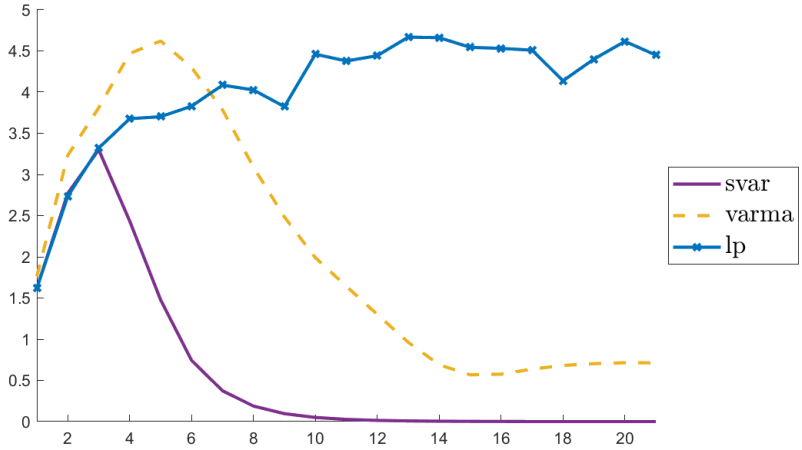
The bias-variance trade-off is a fundamental concept in statistics and machine learning, and it has important implications for the design and evaluation of estimators. For each of the six experiments explained above, I explore this trade-off in the context of fiscal and monetary policy shocks and recursive, observed shock, and IV identification schemes.

Figure 3.1: Recursive G: Mean Bias of Estimators



Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure 3.2: Recursive G: Standard Deviation of Estimators



Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure 3.1 and Figure 3.2 show the absolute bias and standard deviation, respectively, of the VAR, VARMA, and LP estimators at various horizons for the case of fiscal policy shock and recursive identification schemes. The bias and variance are computed by normalizing the units of the response variables by dividing the IRFs by the root mean squared value of the true impulse response function out to horizon 20. These figures illustrate the trade-off between bias and variance in these estimators at different horizons.

Overall, the figures suggest that VAR, VARMA, and LP estimators exhibit similar bias and variance at short horizons, while the LP estimator has a lower bias at longer horizons. Indeed, The LP bias is close to zero at all horizons, yet the variance is high and does not decrease with the horizon. In contrast to the LP estimator, the VAR and VARMA estimators have a higher bias at intermediate horizons, but their variance decreases rapidly as the horizon increases.

From the plots, we can infer that the VARMA(2,1) model exhibits a more extreme bias-variance trade-off than the VAR models. This means that at inter-

mediate horizons, the VARMA model has a lower bias than the VAR model but higher variance at intermediate horizons. This trade-off between bias and variance is also observed between the VAR and LP estimators, with the LP estimator having a lower bias and the VAR estimator having a lower variance, which is a result we can find in the LWP paper. These results suggest that the VARMA model may be a useful estimator in certain contexts where a more extreme trade-off between bias and variance is desired compared to VAR models.

Figures C.1 and Figure C.2 show similar results for the experiment involving monetary policy shocks under the recursive identification scheme. In this case, the bias plot exhibits more jagged behavior but continues to indicate the low degree of bias of the LP method compared to the VAR and VARMA methods, as well as the superiority of VARMA models over VAR models. On the other hand, the variance of the VAR and VARMA models decreases rapidly towards zero, with the fastest decay observed again for the VAR model.

Figure C.7 and C.6 present results for the IV/proxy identification scheme in the fiscal policy shock case. These plots highlight that the external instruments' SVAR-IV procedure faces an even more extreme bias-variance trade-off but in the opposite direction than VARMA models. In this case, the extremely high bias is accompanied by a lower standard deviation. According to LWP, the reason for this is that the VAR-IV is asymptotically biased when the shock is not invertible, and the degree of invertibility is generally low in the DGPs considered. The rest of the results are listed in the appendix section.

3.6 Conclusion and further steps

In this paper, I conduct a small-scale simulation study to investigate the performance of VARMA models in the estimation of IRFs compared to other common methods, such as VAR and LP models. Our simulation design mirrors that of LPW, using dynamic factor models to simulate the dynamics of 207 US macroeconomic time series variables. The main finding is that VARMA models face a trade-off between bias and variance, though not as extreme as the trade-off faced by SVAR-IV methods. These results add to the existing literature on the bias-variance trade-off in the estimation of IRFs for VAR and LP models as presented in LPW and highlight the importance of considering this trade-off when choosing an appropriate estimator for a given application. By understanding the inherent trade-off between bias and variance in these estimators, econometri-

cians can make more informed decisions about which estimator is best suited to a particular problem, ultimately improving the accuracy and reliability of our estimates of IRFs in macroeconomic applications.

This simulation study might inspire several potential avenues for future research. One of these is increasing the number of specifications at the level of LPW, which could provide more robust results but would require intense computational effort. Another interesting extension of this study would be to incorporate the different variations of the estimation methods, such as Bayesian estimations, to see if they improve the estimation of IRFs in VARMA models. Additionally, there are several structural identification schemes in the VARMA literature that could be explored in a large-scale simulation study. Regularization techniques could potentially be implemented to address the dimensionality problem that can arise in models with a large number of variables or high orders of the model.

Overall, there are many opportunities for future research to improve and extend this simulation study, and these avenues of inquiry could provide valuable knowledge into the behavior of VAR and VARMA models.

Appendix A

CHAPTER 1

A.1 Proofs

In this appendix I provide the proofs of the main results. I start by stating a useful lemma from Kristensen and Lee (2019). This lemma is defined for a function $L(\cdot)$ that satisfies: (i) $L(\cdot)$ has a compact support and (ii) for some $C < \infty$, $|L(v) - L(v')| \leq C|v - v'|$, $v, v' \in \mathbb{R}$. The bandwidth re-scaled function is given by $L_b(\cdot) = L(\cdot/b)/b$.

A.1.1 Lemma 1

Lemma 1. *The following holds as $b \to 0$ and $nb \to \infty$:*

(i) Suppose $W_{n,t}(\theta)$ is $ULS(p, q, \Theta)$ with its stationary approximation $W_t^*(\theta|u)$ being L_p continues with $p \geq 1$ and $q > 0$ and Θ compact. For $\mathcal{A}$ defined in assumption 1.2

$$\sup_{\alpha \in \mathcal{A}} \left\| \frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) W_{n,t} \left(D_b(t/n - u)\alpha \right) - \int L(v) \mathbb{E}\left[W_t^*(D(v)\alpha|u) \right] dv \right\| = o_p(1)$$

(ii) Suppose $W_{n,t}(\theta)$ is $ULS(p, q, \Theta)$ with its stationary approximation $W_t^*(\theta|u)$ being L_p continues with $p \geq 1$ and $q > 0$ and Θ compact. Moreover,

for each $u \in [0,1]$, there exist a measurable function $H(u,.)$ such that $W_t^*(\theta|u) = H(u, \mathcal{F}_t)$ and $\delta_q^{W^*}(k) \coloneqq \sup_{u\in[0,1]} \delta_q^{W^*(\theta|u)}(k)$ satisfies that $\triangle_{0,q}^{W^*} \coloneqq \sum_{k=0}^{\infty} \delta_q^{W^*}(k) < \infty$. Then, it holds that:

$$\sqrt{\frac{b}{n}} \sum_{t=1}^{n} L_b(t/n - u) \left\{ W_{n,t}\left(\theta(t/n)\right) - \mathbb{E}\left[W_{n,t}\left(\theta(t/n)\right)\right]\right\} \xrightarrow{d} N\left(0, \int L^2(v)dv\Sigma(u)\right).$$

where $\Sigma(u) = \sum_{k\in\mathbb{Z}} Cov(W_0^*(\theta(u)|u), W_k^*(\theta(u)|u))$

Proof. See Lemma 1 in Kristensen and Lee (2019) for Lemma 1.(i). In order to prove Lemma 1.(ii) I apply theorem 2.10 in Dahlhaus et al. (2019). In order to do so, let's verify whether the conditions in theorem 2.10 in Dahlhaus et al. (2019) hold:

1. Assumption 2.6 in Dahlhaus et al. (2019) regarding the kernel function $K : \mathbb{R} \to \mathbb{R}$ is met by assumption 1.(i). The setup guarantees that the kernel function has to be a bounded function with bounded variation B_K and with compact support $[-1,1]$ satisfying $\int K dx = 1$.[1]

2. Assumption $2.1(S1)$ in Dahlhaus et al. (2019) establish bounds between the stationary approximation and the non-stationary process and fits on assumption 1.(iv) joint with definition 1 as is shown in Vogt (2012). Formally, following definition of the uniformly locally stationary of the process $W_{n,t}(\theta)$ and $p, q > 0$. Take $u = \frac{t}{n}$ and $q \leq 1$:

$$\mathbb{E}\left[\left\| W_{n,t}(\theta) - W_t^*(\theta|\frac{t}{n})\right\|^p\right]^{1/p} \leq \mathbb{E}\left[\sup_{\theta\in\Theta}\left\| W_{n,t}(\theta) - W_t^*(\theta|\frac{t}{n})\right\|^p\right]^{1/p} \leq Cn^{-q}$$

which is the approximation error by replacing the non-stationary process with the stationary approximation. Now, for $v \in [0,1]$:

[1] Dahlhaus (2012) explains the equivalent of the support $[-1,1]$ and $[-\frac{1}{2}, \frac{1}{2}]$ for the kernel.

$$\mathbb{E}\left[\|W_{n,t}(\theta) - W_t^*(\theta|v)\|^p\right]^{1/p} \leq \mathbb{E}\left[\sup_{\theta\in\Theta}\|W_{n,t}(\theta) - W_t^*(\theta|v)\|^p\right]^{1/p} \leq C\left(\left|\frac{t}{n} - v\right|^q + \frac{1}{n^q}\right)$$

By Minkowski inequality:

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|W_{n,t}(\theta) - W_t^*(\theta|v)\|^p\right]^{1/p} \leq \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|W_{n,t}(\theta) - W_t^*(\theta|\frac{t}{n})\right\|^p\right]^{1/p} + \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|W_t^*(\theta|\frac{t}{n}) - W_t^*(\theta|v)\right\|^p\right]^{1/p}$$

with the previous results we can verify that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|W_t^*(\theta|\frac{t}{n}) - W_t^*(\theta|v)\right\|^p\right]^{1/p} \leq C\left(\left|\frac{t}{n} - v\right|\right)^q$$

3. Assumption 2.1 (S2) in Dahlhaus et al. (2019) is met by assumption 1.(iv) and definition 2 regarding $L_p - continuous$ w.r.t θ. As Kristensen and Lee (2019) point out, the definition of $L_p - continuous$ is weaker than almost sure continuity which is required as a condition in Dahlhaus et al. (2019).

4. Finally, Assumption 2.3 (M1) is considered in assumption 2.(2).

Notice that in our case, in assumption 2.(2), $g_t^*(\theta|u)$ is a vector of $q \times 1$ and that theorem 2.10 in Dahlhaus et al. (2019) is stated for scalar random variables. However, Lemma 1.(ii) can also be considered for a random vector by applying Cramer-Wold device, theorem 2.10 in Dahlhaus et al. (2019) and considering that a linear combinations of the elements of $g_t^*(\theta|u)$ retains the property of finite sum of the functional dependence measure.

□

A.1.2 Theorem 1

Proof of Theorem 1. I verify the following conditions: (i) $Q^*(\alpha|u)$ is uniquely minimized at $\alpha^* = (\theta(u)', 0', ..., 0')'$; (ii) Θ is compact; (iii) $Q^*(\alpha|u)$ is continuous in α; (iv) $Q_n(\alpha|u)$ converges uniformly in probability to $Q^*(\alpha|u)$, where

$Q^*(\alpha|u) = g^*(\alpha|u)'\Omega(u)g^*(\alpha|u)$ with

$$g^*(\alpha|u) = \begin{bmatrix} \int K(v)\mathbb{E}\left[g^*\left(v_t, D(v)\alpha|u\right)\right]dv \\ \int K(v)v\mathbb{E}\left[g^*\left(v_t, D(v)\alpha|u\right)\right]dv \\ \vdots \\ \int K(v)v^m\mathbb{E}\left[g^*\left(v_t, D(v)\alpha|u\right)\right]dv \end{bmatrix}.$$

(i) For any $\alpha = (\alpha_1', ..., \alpha_{m+1}')'$ with $\alpha_i \neq 0$ for $i \geq 2$ $D(v)\alpha$ is non-constant almost everywhere, therefore for any $\alpha \neq \alpha^* = (\theta(u)', 0', \ldots, 0')'$ we have $D(v)\alpha \neq D(v)\alpha^* = \theta(u)$ for almost all $v \in [0,1]$ and so by assumption 1.3 and 1.4.(iii) $\Omega(u)\mathbb{E}[g_t^*(D(v)\alpha|u)] \neq 0 = \Omega(u)\mathbb{E}[g_t^*(D(v)\alpha^*|u)] = \Omega(u)\mathbb{E}[g_t^*(\theta(u)|u)]$ for almost every v. Since $K(\cdot) \geq 0$ by assumption 1.1 this in turn implies that $g^*(\alpha|u) \neq 0$ and therefore $Q^*(\alpha|u) \neq 0$. It means that $Q^*(\alpha|u) > Q^*(\alpha^*|u) = 0$ by assumption 1.4.(iii), then α^* uniquely minimizes $Q^*(\theta|u)$.
(ii) is imposed by assumption 1.2.
(iii) is satisfied thanks to assumption 1.4.(ii) and the dominated convergence theorem. Specifically, since $\alpha \mapsto g_t^*(D(v)\alpha|u)$ is L_1 continuous, we have that $\mathbb{E}[\|g_t^*(D(v)\alpha)\|] < \infty$. Take a sequence $\alpha_n \to \alpha$, then

$$\lim_{n\to\infty} \int K(v)v^s\mathbb{E}\left[g_t^*\left(D(v)\alpha_n|u\right)\right] = \int K(v)v^s\mathbb{E}\left[g_t^*\left(D(v)\alpha|u\right)\right],$$

for $s = 0, 1, ..., m$, by the dominated convergence theorem. Therefore, we have that:

$$\lim_{n\to\infty} Q^*(\alpha_n|u) = \lim_{n\to\infty} g^*(\alpha_n|u)'W(u)g^*(\alpha_n|u) = g^*(\alpha|u)'W(u)g^*(\alpha|u) = Q^*(\alpha|u)$$

which show the continuity of $\alpha \mapsto Q^*(\alpha|u)$.
(iv) First, consider $L(v) = K(v)v^s$ for $s = 0, 1, ..., m$. By assumption 1.1 we have

$$\begin{aligned} |L(v) - L(v')| = |K(v)v^s - K(v')v'^s| &\leq |K(v)v^s - K(v')v^s| + |K(v')v^s - K(v')v'^s| \\ &\leq |K(v) - K(v')||v^s| + |K(v')||(v^s - v'^s)| \\ &\leq C|(v - v')||v^s| + |K(v')||(v^s - v'^s)| \\ &\leq C'|(v - v')| \end{aligned}$$

then $L(v) = K(v)v^s$ satisfies the condition for applying Lemma 1 with $L_b(\cdot) = [K(\cdot/b)/b](\cdot/b)^s$. By assumption 1.4.(i) $g(v_{n,t}, \theta)$ is $ULS(p, q, \Theta)$ for $p \geq 1$,

$q > 0$ with L_1 stationary approximation $g_t^*(\theta|u)$ and Θ being compact by assumption 1.2. Then, for $\mathcal{A}$ defined in assumption 1.2, all conditions are satisfied, so I can apply lemma 1.(i) for $s = 0$

$$\sup_{\alpha\in\mathcal{A}} \left\| \frac{1}{n}\sum_{t=1}^{n} L_b(t/n-u)g(v_{n,t}, D_b(t/n-u)\alpha) - \int L(v)\mathbb{E}\left[g_t^*\left(D(v)\alpha|u\right)\right]dv \right\| = o_p(1)\,.$$

and for $s > 0$ we have

$$\sup_{\alpha\in\mathcal{A}} \left\| \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u)((t/n-u)/b)^s g(v_{n,t}, D_b(t/n-u)\alpha) - \int K(v)v^s\mathbb{E}\left[g_t^*\left(D(v)\alpha|u\right)\right]dv \right\| = o_p(1)\,.$$

From the previous results and assumption 1.4, I get that:

$$\sup_{\alpha\in\mathbb{A}} |Q_n(\alpha|u) - Q^*(\alpha|u)| = o_p(1).$$

Therefore, conditions (i)-(iv) in theorem 2.1 in Newey and McFadden (1994b) are satisfied which imply that $\hat{\alpha}(u) \to \alpha^*(u)$. □

A.1.3 Theorem 2

Proof of Theorem 2. Recall that $s(v_{n,t},\theta) = \frac{\partial g(v_{n,t},\theta)}{\partial\theta'}$ and let $G_n(\alpha|u) = \frac{\partial g_n(\alpha|u)}{\partial\alpha'}$ such that

$$G_n(\alpha|u) = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u) \begin{bmatrix} s(v_{n,t}, D_b(t/n-u)\alpha) \\ s(v_{n,t}, D_b(t/n-u)\alpha)(\frac{t/n-u}{b})^1 \\ \vdots \\ s(v_{n,t}, D_b(t/n-u)\alpha)(\frac{t/n-u}{b})^m \end{bmatrix} D_b(t/n-u)\,.$$

At the optimum $\hat{\alpha} = \arg\min_{\alpha\in\mathcal{A}} Q_n(\alpha|u)$ the first order conditions of $Q_n(\alpha|u)$ with respect to α are given by

$$0 = G_n(\hat{\alpha}|u)'\Omega_n(u)g_n(\hat{\alpha}|u)\,. \tag{A.1}$$

From theorem 1 we know that $\hat{\alpha} \xrightarrow{p} (\theta(u)', 0', \ldots, 0')'$ which is an interior point of $\mathcal{A}$ by Assumption 2.1.(ii); then with probability approaching one we have that $\hat{\alpha}$ is also an interior point of $\mathcal{A}$. This implies that for n sufficiently large we can apply the mean value theorem to $g_n(\hat{\alpha}|u)$ and get

$$g_n(\hat{\alpha}|u) = g_n(\alpha|u) + G_n(\tilde{\alpha}|u)(\hat{\alpha} - \alpha) \tag{A.2}$$

where $\tilde{\alpha}$ lies between $\hat{\alpha}$ and α. Pre-multiplying (A.2) by $G_n(\hat{\alpha}|u)'\Omega_n(u)$ and using (A.1) gives

$$\begin{aligned} 0 = G_n(\hat{\alpha}|u)'\Omega_n(u)g_n(\hat{\alpha}|u) = G_n(\hat{\alpha}|u)'\Omega_n(u)g_n(\alpha|u) + \\ G_n(\hat{\alpha}|u)'\Omega_n(u)G_n(\tilde{\alpha}|u)(\hat{\alpha} - \alpha) \, . \end{aligned}$$

Recall, $\theta_u^*(t/n) = D(t/n - u)\beta = D_b(t/n - u)$, such that (slightly abusing notation) $g_n(\alpha|u) = g_n(\theta_u^*(t/n)|u)$, which are the sample moments evaluated at the approximating parameter $\theta_u^*(t/n)$. I also define

$$g_n(\theta(t/n)|u) = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n - u) \begin{bmatrix} g(v_{n,t}, \theta(t/n)) \\ g(v_{n,t}, \theta(t/n))(\frac{t/n-u}{b})^1 \\ \vdots \\ g(v_{n,t}, \theta(t/n))(\frac{t/n-u}{b})^m \end{bmatrix} ,$$

This allows us to decompose the moment function into bias and variance components by adding and subtracting $g_n(\theta(t/n)|u)$ from the first order condition.

$$\begin{aligned} 0 =& G_n(\hat{\alpha}|u)\Omega_n(u) \left(g_n(\theta_u^*(t/n)|u) - g_n(\theta(t/n)|u) + g_n(\theta(t/n)|u)\right) \\ &+ G_n(\hat{\alpha}|u)'\Omega_n(u)G_n(\tilde{\alpha}|u)(\hat{\alpha} - \alpha) \end{aligned} \tag{A.3}$$

I define

$$\begin{aligned} \mathbb{B}_n(u) =& G_n(\hat{\alpha}|u)\Omega_n(u)b_n(u) \\ \mathbb{S}_n(u) =& G_n(\hat{\alpha}|u)\Omega_n(u)g_n(\theta(t/n)|u) \\ \mathbb{H}_n(u) =& G_n(\hat{\alpha}|u)'\Omega_n(u)G_n(\tilde{\alpha}|u) \end{aligned}$$

where $b_n(u) = g_n(\theta_u^*(t/n)|u) - g_n(\theta(t/n)|u)$. Rearranging terms and multiplying (B.8) by $\sqrt{nb}$ gives

$$0 = \sqrt{nb}\mathbb{S}_T(u) + \mathbb{H}_n(u)\sqrt{nb}\left(\hat{\alpha} - \alpha + \mathbb{H}_n(u)^{-1}\mathbb{B}_n(u)\right) \, .$$

To work out the limit of this expression, we need to prove the following:

(i) $\sqrt{nb}g_n(\theta(t/n)|u) \xrightarrow{d} N(0, \mathbb{K} \otimes \Sigma_\omega(u))$

(ii) $G_n(\hat{\alpha}|u) \xrightarrow{p} G_0(u)$ and $G_n(\tilde{\alpha}|u) \xrightarrow{p} G_0(u)$, where $G_0(u) = (I_{m+1} \otimes \Sigma_s(u)) \times \mathbb{J}$

(iii) $b_n(u) \xrightarrow{p} -b^{m+1}\left(\mathbb{M} \otimes \Sigma_\omega(u)\frac{\theta^{(m+1)}(u)}{(m+1)!} + o_p(1)\right) = -b^{m+1}\left(\mathbb{B}(u) + o_p(1)\right)$

where:

$$\mathbb{K} = \begin{bmatrix} \mathbb{K}_0^2 & \mathbb{K}_1^2 & \dots & \mathbb{K}_m^2 \\ \mathbb{K}_1^2 & \mathbb{K}_2^2 & \dots & \mathbb{K}_{m+1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{K}_m^2 & \mathbb{K}_{m+1}^2 & \dots & \mathbb{K}_{2m}^2 \end{bmatrix} \in \mathbb{R}^{(m+1)\times(m+1)},$$

$\mathbb{K}_i^2 = \int K^2(v)v^i dv$, $\mathbb{M} = diag\{\mathbb{K}_0^1, \mathbb{K}_1^1, \dots, \mathbb{K}_m^1\}$, and $\mathbb{J} = \begin{bmatrix}\mu_0' & \mu_1' & \dots & \mu_m'\end{bmatrix}'$ with $\mathbb{K}_i^1 = \int K(v)v^{m+1+i}dv$ and $\mu_i = \int K(v)v^i D(v)dv$.

For results (i), first note that:

$$\sqrt{nb}g_n(\theta|u) = \sqrt{\frac{b}{n}}\sum_{t=1}^{n} V(t/n - u) \otimes g(v_{n,t}, \theta)$$

where $V(t/n-u) = K_b(t/n-u)(1, (\frac{t/n-u}{b})^1, \dots, (\frac{t/n-u}{b})^m)'$ and note that $\int V(v)V(v)'\mathrm{d}v = \mathbb{K}$.

By Assumptions 1.1 and application of lemma 1.(ii) and Cramer-Wold device we get $\sqrt{nb}g_n(\theta(t/n)|u) \xrightarrow{d} N(0, \mathbb{K} \otimes \Sigma(u))$ where

$$\Sigma(u) = \sum_{k\in\mathbb{Z}} Cov(g_0^*(\theta(u)|u), g_k^*(\theta(u)|u)).$$

Now, let's go forward with result (ii). We know that:

$$G_n(\alpha|u) = \frac{1}{n}\sum_{t=1}^{n} V(t/n-u) \otimes s(v_{n,t}, D_b(t/n-u)\alpha)D_b(t/n-u)$$

We want to apply lemma 1.(i) to $s(v_t, D_b(t/n-u)\alpha)$. Let's verify whether the conditions are satisfied in this case. By similar argument in theorem 1 we know that $L(v) = K(v)v^s D(v)$ for $s = 0, 1, ..., m$ satisfies condition in lemma 1.(i). Assumption 2 guarantees that $s(v_{n,t}, \theta)$ is $ULS\ (p, q, \{\theta : ||\theta - \theta(u)|| < \epsilon)\}$ for

some $p \geq 1$ and $q, \epsilon > 0$ with L_1 continuous stationary approximation $s_t^*(\theta|u)$ with $\Sigma_s(u) = E(s_t^*(\theta(u)|u))$. All conditions in lemma 1.(i) are met. Then, with $\mathcal{B}(\epsilon) = \{\alpha : ||\alpha - \alpha^*|| < \epsilon\}$ for some $\epsilon > 0$:

$$\sup_{\alpha \in \mathcal{B}(\epsilon)} ||\frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u)s(v_{n,t}, D_b(t/n-u)\alpha)(\frac{t/n-u}{b})^s D_b(t/n-u) - \int K(v)v^s \mathbb{E}[s_t^*(D(v)\alpha|u)]D(v)dv|| = o_p(1)$$

for $s = 0, 1, ..., m$, where $\mathbb{E}[s_t^*(\theta|u)]$ is continuous w.r.t θ by assumption 2. Then, given that $\hat{\alpha}(u) \to \alpha^*(u) = (\theta(u)', \mathbf{0}', ..., \mathbf{0}')'$ we can apply the previous results and obtain:

$$\sup_{\alpha \in \mathcal{B}(\epsilon)} ||\frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u)s(v_{n,t}, D_b(t/n-u)\hat{\alpha}(u)))(\frac{t/n-u}{b})^s D_b(t/n-u) - \Sigma_s(u)\int K(v)v^s D(v)dv|| = o_p(1)$$

If we stack the previous results for all $s = 0, 1, ..., m$ we obtain:

$$G_n(\hat{\alpha}(u)|u) \to (I_{m+1} \otimes \Sigma(u)) \times \mathbb{J}$$

with $\mathbb{J} = \begin{bmatrix}\mu_0' & \mu_1' & \dots & \mu_m'\end{bmatrix}'$ where $\mu_i = \int K(v)v^i D(v)dv \in \mathbb{R}^{p \times p(m+1)}$. Moreover, we know that $\tilde{\alpha}(u) \to \alpha^*(u)$ because $\tilde{\alpha}(u)$ is on the line between $\hat{\alpha}(u)$ and $\alpha^*(u)$. Thus we can apply the same arguments and get:

$$G_n(\tilde{\alpha}(u)|u) \to (I_{m+1} \otimes \Sigma(u)) \times \mathbb{J}$$

Next, we continue with result (iii). Remember that

$$b_n(u) = g_n(\theta_u^*(t/n)|u) - g_n(\theta(t/n)|u)$$

$$b_n(u) = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u) \begin{bmatrix} g(v_{n,t}, \theta_u^*(t/n)) - g(v_{n,t}, \theta(t/n)) \\ \Big(g(v_{n,t}, \theta_u^*(t/n)) - g(v_{n,t}, \theta(t/n))\Big)(\frac{t/n-u}{b}) \\ \vdots \\ \Big(g(v_{n,t}, \theta_u^*(t/n)) - g(v_{n,t}, \theta(t/n))\Big)(\frac{t/n-u}{b})^m \end{bmatrix}$$

We can apply the mean-value theorem twice and obtain that for some $\bar{\theta}(t/n)$ on the line between $\theta_u^*(t/n)$ and $\theta(t/n)$ and some $u' \in [t/n, u]$:

$$\begin{aligned} g(v_{n,t}, \theta_u^*(t/n)) - g(v_{n,t}, \theta(t/n)) =& s(v_{n,t}, \bar{\theta}(t/n))(\theta_u^*(t/n) - \theta(t/n)) \\ =& - s(v_{n,t}, \bar{\theta}(t/n)) \frac{\theta^{(m+1)}(u')}{(m+1)!}(t/n - u)^{m+1} \end{aligned}$$

Let's add and subtract $s(v_{n,t}, \theta(t/n)) \frac{\theta^{(m+1)}(t/n)}{(m+1)!}(t/n-u)^{m+1}$:

$$\begin{aligned} =& -s(v_{n,t}, \theta(t/n)) \frac{\theta^{(m+1)}(t/n)}{(m+1)!}(t/n-u)^{m+1} + \\ & \Big\{ s(v_{n,t}, \theta(t/n))\theta^{(m+1)}(t/n) - s(v_{n,t}, \bar{\theta}(t/n))\theta^{(m+1)}(u') \Big\} \frac{(t/n-u)^{m+1}}{(m+1)!} \end{aligned} \tag{A.4}$$

By assumption 2, the first term is ULS and by applying lemma 1.(i) we get:

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u) s(v_{n,t}, \theta(t/n)) \frac{\theta^{(m+1)}(t/n)}{(m+1)!}(t/n-u)^{m+1}(\frac{t/n-u}{b})^i \\ \to b^{m+1}\mathbb{K}_i^1 \Sigma_s(u) \frac{\theta^{(m+1)}(u)}{(m+1)!} \end{aligned}$$

where $\mathbb{K}_i^1 = \int K(v) v^{m+1+i} dv$. For the remaining term, notice that $|t/n - u| < Cb$ and $||\bar{\theta}(t/n) - \theta(t/n)|| \leq ||\theta_u^*(t/n) - \theta(t/n)||$ because $\bar{\theta}(t/n)$ is on the line between $\theta_u^*(t/n)$ and $\theta(t/n)$. Using these inequalities and consider the Taylor approximation in 1.2 we get $||\bar{\theta}(t/n) - \theta(t/n)|| \leq \tilde{C} b^{m+1}$. We can use the ULS properties of $s(v_{n,t}, \theta)$ in assumption 2 and establish:

$$\begin{aligned} & \sup_{n,t} \mathbb{E}[||s(v_{n,t}, \theta(t/n)) - s(v_{n,t}, \bar{\theta}(t/n))||] \\ \leq & \mathbb{E}[\sup_{n,t} ||s(v_{n,t}, \theta(t/n)) - s(v_{n,t}, \bar{\theta}(t/n))||] \leq \\ C\Big(b^q |\frac{t/n-u}{b}|^q + \frac{1}{n^q} \Big) & + \sup_{||\theta - \theta'|| \leq \tilde{C} b^{m+1}} \mathbb{E}[||s_t^*(\theta|u) - s_t^*(\theta'|u)|| \to 0 \end{aligned}$$

as $n \to \infty$. In the same way, we can use the uniform continuity of $\theta^{(m+1)}(.)$ and obtain that $\sup_{n,t} ||\theta^{(m+1)}(t/n) - \theta^{(m+1)}(t/n)(u')|| \to 0$ as $n \to \infty$ because

$u' \in [t/n), u]$. Therefore, these results guarantee that the second term in A.4 is $o_p(1)$. Thus,

$$b_T(u) \to -b^{m+1}\Big(\mathbb{M} \otimes \Sigma_s(u)\frac{\theta^{(m+1)}(u)}{(m+1)!} + o_p(1)\Big)$$

where $\mathbb{M} = diag\{\mathbb{K}_0^1, \mathbb{K}_1^1, \ldots, \mathbb{K}_m^1\}$. The result of theorem 2 comes after applying Slusky's theorem to results (i), (ii) and (iii). □

A.2 Tables

Table A.1: TV-MA(1) model: MADE

		TV_1		TV_2		TV_3	
	u	θ_t	σ_t^2	θ_t	σ_t^2	θ_t	σ_t^2
T=500	0.10	0.1760	0.2218	0.1229	0.1125	0.1419	0.1405
	0.50	0.1572	0.1638	0.1501	0.1377	0.1778	0.1929
	0.90	0.1212	0.1093	0.1847	0.2139	0.1842	0.2229
T=1000	0.10	0.1542	0.1756	0.0905	0.0815	0.1068	0.1018
	0.50	0.1132	0.1201	0.1073	0.1002	0.1403	0.1489
	0.90	0.0875	0.0803	0.1580	0.1701	0.1601	0.1781

The table shows the mean absolute deviation error (MADE) for each of the parameters θ_t, σ_t at different points in time ($u = 0.10, 0.50, 0.90$) for two sample sizes $T = 500, 1000$. There are three different functional forms for the true parameters: TV_1 the coefficients take a cosine shape; TV_2 the coefficients are linear trend with a break at $u = 0.50$; and TV_3 the coefficients take a square root function.

Table A.2: TV-MA(1) model: RMSE

		TV_1		TV_2		TV_3	
	u	θ_t	σ_t^2	θ_t	σ_t^2	θ_t	σ_t^2
T=500	0.10	0.2127	0.2688	0.1638	0.1397	0.1868	0.1738
	0.50	0.2028	0.2011	0.1969	0.1687	0.2180	0.2354
	0.90	0.1644	0.1355	0.2206	0.2580	0.2191	0.2695
T=1000	0.10	0.1914	0.2157	0.1197	0.1017	0.1451	0.1269
	0.50	0.1500	0.1508	0.1456	0.1258	0.1809	0.1851
	0.90	0.1172	0.1028	0.1962	0.2093	0.1985	0.2193

The table shows the root of the mean squared errors (RMSE) for each of parameters θ_t, σ_t at different points in time ($u = 0.10, 0.50, 0.90$) for two sample sizes $T = 500, 1000$. There are three different functional forms for the true parameters: TV_1 the coefficients take a cosine shape; TV_2 the coefficients are linear trend with a break at $u = 0.50$; and TV_3 the coefficients take a square root function.

Table A.3: TV-MA(1) model: Coverage

		TV_1		TV_2		TV_3	
	u	θ_t	σ^2	θ_t	σ^2	θ_t	σ^2
T=500	0.10	0.9210	0.9730	0.9530	0.9120	0.9420	0.9320
	0.50	0.9430	0.9400	0.9460	0.9340	0.9380	0.9590
	0.90	0.9660	0.9390	0.9290	0.9640	0.9300	0.9670
T=1000	0.10	0.9240	0.9730	0.9490	0.9400	0.9450	0.9480
	0.50	0.9490	0.9500	0.9510	0.9320	0.9380	0.9610
	0.90	0.9640	0.9430	0.9340	0.9830	0.9390	0.9820

The table shows the coverage for each of the parameters at different points in time ($u = 0.10, 0.50, 0.90$) for two sample sizes $T = 500, 1000$. The coverage is defined as the proportion of time where a 95% confidence interval contains the true value. There are three different functional forms for the true parameters: TV_1 the coefficients take a cosine shape; TV_2 the coefficients are linear trend with a break at $u = 0.50$; and TV_3 the coefficients take a square root function.

Table A.4: TV-ARCH model: α_t

		MADE			RMSE		
	u	TV_1	TV_2	TV_3	TV_1	TV_2	TV_3
T=500	0.10	0.1395	0.1004	0.1111	0.1642	0.1143	0.1247
	0.50	0.1012	0.1069	0.1231	0.1151	0.1228	0.1478
	0.90	0.1380	0.1799	0.1557	0.1640	0.2146	0.1866
T=1000	0.10	0.1190	0.0868	0.0970	0.1454	0.1008	0.1124
	0.50	0.0915	0.0982	0.1148	0.1051	0.1164	0.1408
	0.90	0.1249	0.1533	0.1367	0.1506	0.1880	0.1671
T=2000	0.10	0.1033	0.0792	0.0881	0.1298	0.0909	0.1040
	0.50	0.0793	0.0855	0.0979	0.0933	0.1023	0.1231
	0.90	0.1062	0.1222	0.1123	0.1324	0.1531	0.1410
T=4000	0.10	0.0892	0.0718	0.0799	0.1132	0.0826	0.0952
	0.50	0.0718	0.0772	0.0863	0.0857	0.0945	0.1113
	0.90	0.0867	0.0949	0.0895	0.1107	0.1219	0.1177

The table shows the mean absolute deviation error (MADE) and the root of the mean squared errors (RMSE) for the parameters α_t at different points in time ($u = 0.10, 0.50, 0.90$) for two sample sizes $T = 500, 1000, 2000, 4000$. There are three different functional forms for the true parameters: TV_1 the coefficients take a cosine shape; TV_2 the coefficients are linear trend with a break at $u = 0.50$; and TV_3 the coefficients take a square root function.

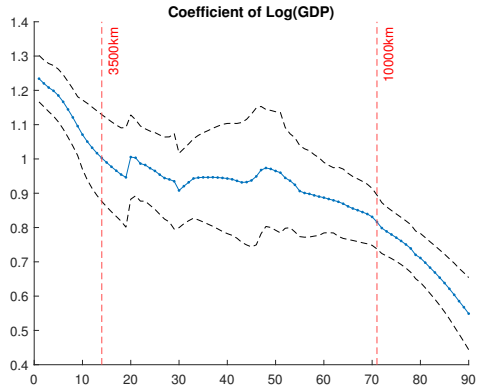
Table A.5: TV-ARCH(1): Coverage

	u	TV_1	TV_2	TV_3
T=500	0.10	0.8990	0.8670	0.8930
	0.50	0.8680	0.8750	0.8800
	0.90	0.8940	0.7770	0.8470
T=1000	0.10	0.9060	0.8930	0.9200
	0.50	0.9110	0.9230	0.9020
	0.90	0.8800	0.7380	0.8220
T=2000	0.10	0.9090	0.9120	0.9310
	0.50	0.9310	0.9390	0.9270
	0.90	0.8990	0.7330	0.8360
T=4000	0.10	0.9010	0.9150	0.9460
	0.50	0.9380	0.9350	0.9000
	0.90	0.8840	0.7270	0.8020

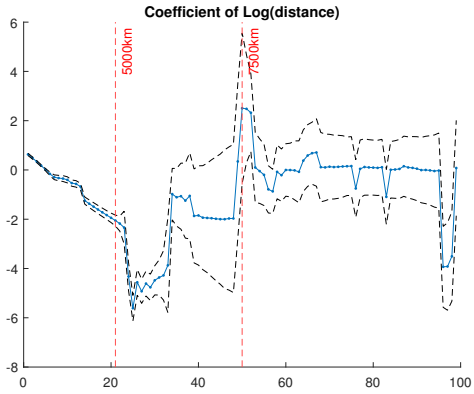
The table shows the coverage for the parameters α_t at different points in time ($u = 0.10, 0.50, 0.90$) for two sample sizes $T = 500, 1000, 2000, 4000$. The coverage is defined as the proportion of time where a 95% confidence interval contains the true value. There are three different functional forms for the true parameters: TV_1 the coefficients take a cosine shape; TV_2 the coefficients are linear trend with a break at $u = 0.50$; and TV_3 the coefficients take a square root function.

A.3 Figures

Figure A.1: Gravity Model with "Distance" Varying Coefficients



(a)



(b)

Note: The Plots display the distance-varying coefficient of a gravity model for the US and 100 top partners by implementing the nonparametric GMM local linear estimator. A 90% confidence set is computed with HAC standard errors. Panel (a) shows results for GDP (in logs) of the partner and Panel (b) depicts results for the distance variable (in logs).

Appendix B

CHAPTER 2

B.1 Proof of Lemma 1

Proof. Consider first the case where:

$$\hat{\beta}_h^f - \hat{\beta} \xrightarrow{d} N(0, MSE(\hat{\beta}_h^f))$$

The optimization explained in section 2.5 determines that $\hat{\beta}_h^f$ is the best fit to $\hat{\beta}$ and this difference is the approximation error. Therefore we can say:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f - \hat{\beta}}{\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}} \leq z_\alpha\Big) = 1 - \alpha$$

By adding $\hat{\beta} - \beta^\dagger$ taking advantage that $\hat{\beta}$ is a consistent estimator of $\beta^\dagger$ as $n \to \infty$:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f - \hat{\beta}}{\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}} + \frac{\hat{\beta} - \beta^\dagger}{\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}} \leq z_\alpha\Big) = 1 - \alpha$$

As a result, we have proven our desired result.

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f - \beta^\dagger}{\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}} \leq z_\alpha\Big) = 1 - \alpha$$

Now consider the case where

$$\hat{\beta}_h^f - \hat{\beta} \xrightarrow{d} N(Abias(\hat{\beta}_h^f), Avar(\hat{\beta}_h^f))$$

where $MSE(\hat{\beta}_h^f) = Abias(\hat{\beta}_h^f)^2 + Avar(\hat{\beta}_h^f)$. In this case, the following is true:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f - \hat{\beta} - Abias(\hat{\beta}_h^f)}{\sqrt{Avar(\hat{\beta}_h^f)}} \leq z_\alpha\Big) = 1 - \alpha$$

$$\lim_{n\to\infty} Pr\Big(\hat{\beta}_h^f - \hat{\beta} \leq Abias(\hat{\beta}_h^f) + z_\alpha\sqrt{Avar(\hat{\beta}_h^f)}\Big) = 1 - \alpha$$

Let's compare this expression with the one in the lemma:

$$\lim_{n\to\infty} Pr\Big(\hat{\beta}_h^f - \hat{\beta} \leq z_\alpha\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}\Big)$$

Notice that for high levels of bias relative to the variance, the expression in the lemma is bigger which is the most likely case with one or few basis functions. Then we can affirm that:

$$\lim_{n\to\infty} Pr\Big(\hat{\beta}_h^f - \hat{\beta} \leq z_\alpha\sqrt{\hat{MSE}(\hat{\beta}^f)_{hh}}\Big) \geq 1 - \alpha$$

□

B.2 Proof of Lemma 2

Proof. Following similar strategy than in lemma 1. Consider first the case where:

$$\hat{\beta}_h^f(u) - \hat{\beta}(u) \xrightarrow{d} N\Big(0, MSE(\hat{\beta}_h^f(u))\Big)$$

The optimization explained in section 2.7.2 determines that $\hat{\beta}_h^f(u)$ is the best fit to $\hat{\beta}(u)$ and this difference is the local approximation error. Therefore we can say:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f(u) - \hat{\beta}(u)}{\sqrt{\hat{MSE}(\hat{\beta}^f(u))_{hh}}} \leq z_\alpha\Big) = 1 - \alpha$$

By adding $\hat{\beta}(u) - \beta(u)$ taking advantage that $\hat{\beta}(u)$ is a consistent estimator of $\beta(u)$ as $n \to \infty$:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f(u) - \hat{\beta}(u)}{\sqrt{M\hat{S}E(\hat{\beta}^f(u))_{hh}}} + \frac{\hat{\beta} - \beta(u)}{\sqrt{M\hat{S}E(\hat{\beta}^f)_{hh}(u)}} \leq z_\alpha\Big) = 1 - \alpha$$

As a result, we obtain:

$$\lim_{n\to\infty} Pr\Big(\frac{\hat{\beta}_h^f(u) - \beta(u)}{\sqrt{M\hat{S}E(\hat{\beta}^f)_{hh}}} \leq z_\alpha\Big) = 1 - \alpha$$

The proof for the other case follows the same logic as the comparison in the proof of lemma 1, so it has been omitted here.

□

B.3 GMM Local Linear Estimator for Linear Models

Let's consider a h-step predictive regression model with endogenous regressors and time-varying coefficients:

$$y_{t+h} = x_t' \delta^h(t/n) + \epsilon_{t+h}^h \tag{B.1}$$

for $h = 0, 1, 2, ...H$, where x_t and $\delta^h(t/n)$ are $d \times 1$ vectors, and y_{t+h} is an scalar. Let's assume that $\delta^h(t/n)$ can be approximated by a linear function at any fixed time point $u \in [0, 1]$ as follows:

$$\delta^h(t/n) = a_u^h + b_u^h(t/n - u)$$

where $\delta^h(u) = a_u^h$ and $\delta^{h'}(u) = b_u^h$. Therefore, the model can be written around u as:

$$y_{t+h} = X_{t,u}' \theta_u^h + \epsilon_{t+h}^h \tag{B.2}$$

where $X_{t,u} = [x_t \quad x_t(t/n - u)]$ and $\theta_u^h = [a_u^h \quad b_u^h]$ and $[\delta^h(u) \quad \delta^{h'}(u)]$ are $2d \times 1$ vectors.

We assumed there is available a $m \times 1$ vector of instrumental variables z_t such $\mathbb{E}(\epsilon_{k+h}^h | z_t) = 0$. In the same way than before, let's define the $2m \times 1$

vector $Z_{t,u} = [z_t \quad z_t(t/n-u)]$. Notice that $\mathbb{E}(Z_{t,u}\epsilon^h_{k+h}) = 0$ because of the exogeneity of the instruments.

Let's define an estimator $\hat{\theta}^h_u(\hat{W})$ such that it is obtained by minimizing the following GMM-form loss function:

$$\hat{\theta}^h_u(\hat{W}) = arg\min_{\tilde{\theta}} \Big\{ \Big(\frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) g^h_{t,u}(\tilde{\theta}) \Big)' \hat{W} \Big(\frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) g^h_{t,u}(\tilde{\theta}) \Big) \Big\} \tag{B.3}$$

where $g^h_{t,u}(\theta) = Z_{t,u}\epsilon^h_{t+h} = Z_{t,u}(y_{t+h} - X'_{t,u}\theta)$ which satisfies the moment conditions $\mathbb{E}(Z_{t,u}\epsilon^h_{k+h}) = 0$. The function $k_b(t/n-u)$ is defined as $\frac{k(\frac{u-t/n}{b})}{b}$ where $k(.)$ is a kernel function with bandwidth parameter b which satisfies that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. Moreover, $\hat{W} = H_2^{-1}\hat{W}H_2^{-1}$ such that $\hat{W}$ is a symmetric positive definite $2m \times 2m$ matrix and $H_2 = diag(1,b) \otimes I_{(m)}$ with $I_{(m)}$ as the identity matrix of order m.

We can rewrite the estimator as:

$$\hat{\theta_u}^h(\hat{W}) = arg\min_{\tilde{\theta}} \Big\{ \Big(\frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) Z_{t,u}(y_{t+h} - X'_{t,u}\tilde{\theta}) \Big)' \hat{W} \Big(\frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) Z_{t,u}(y_{t+h} - X'_{t,u}\tilde{\theta}) \Big) \Big\}$$

Let's define:

$$S^{u,h}_{Zy} = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) Z_{t,u} y_{t+h} \tag{B.4}$$

$$S^{u}_{ZX} = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) Z_{t,u} X'_{t,u} \tag{B.5}$$

In matrix form:

$$S^{u,h}_{Zy} = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t y_{t+h} \\ \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t y_{t+h}(t/n-u) \end{bmatrix} \tag{B.6}$$

$$S^{u}_{ZX} = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t x'_t & \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t x'_t(t/n-u) \\ \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t x'_t(t/n-u) & \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u) z_t x'_t(t/n-u)^2 \end{bmatrix} \tag{B.7}$$

Expressing the nonparametric estimator in terms of $S^{u,h}_{Zy}$ and S^{u}_{ZX}:

$$\hat{\theta}_u^h(\hat{W}) = arg\min_{\tilde{\theta}} \left\{ (S_{Zy}^{u,h} - S_{ZX}^u \tilde{\theta})' \hat{W} (S_{Zy}^{u,h} - S_{ZX}^u \tilde{\theta}) \right\}$$

The first order condition for the above minimization problem is:

$$S_{ZX}^{u\prime} \hat{W} S_{Zy}^{u,h} = S_{ZX}^{u\prime} \hat{W} S_{ZX}^u \tilde{\theta}$$

Hence,

$$\hat{\theta_u}^h(\hat{W}) = (S_{ZX}^{u\prime} \hat{W} S_{ZX}^u)^{-1} S_{ZX}^{u\prime} \hat{W} S_{Zy}^{u,h} \tag{B.8}$$

$$\begin{bmatrix} \hat{\delta}^h(u) \\ \hat{\delta}^{h\prime}(u) \end{bmatrix} = (S_{ZX}^{u\prime} \hat{W} S_{ZX}^u)^{-1} S_{ZX}^{u\prime} \hat{W} S_{Zy}^{u,h} \tag{B.9}$$

The local linear estimator for $\delta^h(u)$ is just the first d rows of $\hat{\theta_u}^h(\hat{W})$:

$$\hat{\delta}^h(u) = (e_1' \otimes I_{(d)})(S_{ZX}^{u\prime} \hat{W} S_{ZX}^u)^{-1} S_{ZX}^{u\prime} \hat{W} S_{Zy}^{u,h} \tag{B.10}$$

where $e_1 = [1 \quad 0]'$

B.3.1 Assumptions

Assumption 1. *The set of m instruments, z_t, satisfy:*

$$\mathbb{E}(\epsilon_{t+h}^h | z_t) = 0 \tag{B.11}$$

Assumption 2. *Assume $(x_t', z_t', \epsilon_{t+h}^h)$ is a α-mixing stochastic process such that $\exists\ \delta > 0, E \mid x_t \mid^{2(2+\delta)} < \infty$ and $E \mid z_t \mid^{2(2+\delta)} < \infty$. Additionally, the mixing coefficients satisfy $\alpha(p) = O(p^{-\phi})$ with $\phi = \frac{(2+\delta)(1+\delta)}{\delta}$.*

Assumption 3. *(i) $\delta^h(t/n) : [0,1] \to \mathbb{R}^d$ is a smooth function such that its second order derivative is continuous in $[0,1]$. (ii) $M_{zx}(t/n) = E(z_t x_t') : [0,1] \to \mathbb{R}_d^{m \times d}$ is a smooth function such that its second order derivative is continuous in $[0,1]$ where $\mathbb{R}_d^{m \times d}$ denotes the set of matrices of order $m \times d$ which are of full column rank. (iii) $\Gamma_s(t/n) = cov(z_t \epsilon_{t+h}^h, z_{t+s} \epsilon_{t+h+s}^h)$ is a smooth function such that its second order derivative is continuous in $[0,1]$.*

Assumption 4. *The kernel function $k(.) : [-1,1] \to \mathbb{R}$ is symmetric bounded probability density function which satisfies the Lipschitz condition. Therefore, $\int_{-1}^{1} k(u)du = 1$ and for any j, $\int_{-1}^{1} u^{2j+1} k(u)du = 0$.*

Assumption 5. *$\hat{W}$ is a symmetric positive definite $2m \times 2m$ matrix such that $\hat{W} \xrightarrow{p} W$ with W symmetric positive definite.*

Assumption 6. $nb^{1+\frac{4}{\delta}} \to \infty$.

B.3.2 Asymptotic Theory

The following theorems summarize the main important asymptotic properties of the proposed estimator. Proof of theorem 1 and 2 can be found in Appendix A and B, respectively.

Theorem 1. *Consistency. Under assumptions 1-5, for any $u \in (0,1)$ and $h = 1, 2, ..., H$, as $n \to \infty$ we have:*

$$H_1(\hat{\theta_u}^h(\hat{W}) - \theta_u^h) - bias \to o_p(b^2) + O_p((Tb)^{-1/2}) \tag{B.12}$$

where $bias = \frac{1}{2}b^2\{M(u)'WM(u)\}^{-1}M(u)'W\begin{bmatrix}\mu_2 M_{zx}(u)\delta^{h\prime\prime}(u)\\ \mu_3 M_{zx}(u)\delta^{h\prime\prime}(u)\end{bmatrix}$; $M(u) = \begin{bmatrix}\mu_0 M_{zx}(u) & \mu_1 M_{zx}(u)\\ \mu_1 M_{zx}(u) & \mu_2 M_{zx}(u)\end{bmatrix}$; $\mu_j = \int u^j k(u)du$ *for* $j = 0, 1, 2, 3$; $H_1 = diag(1, b) \otimes I_{(d)}$ *such that $I_{(d)}$ is the identity matrix of order d; and* $\delta^{h\prime\prime}(u) = \frac{\partial^2 \delta^h(u)}{\partial u^2}$. *In the particular case when $d = m$ (the number of instrumental variables is equal to the number of endogenous variables), the expression for the $bias$ term simplifies to* $bias = \frac{1}{2}b^2\begin{bmatrix}\mu_2\delta^{h\prime\prime}(u)\\ 0\end{bmatrix}$ *which is similar to the bias term in Cai (2007) and Chen (2015).*

Theorem 2. *Asymptotic Normality. Under assumptions 1-6, for any $u \in (0,1)$ and $h = 1, 2, ..., H$, as $n \to \infty$ we have:*

$$\sqrt{nb}[H_1(\hat{\theta_u}^h(\hat{W}) - \theta_u^h) - bias + o_p(b^2)] \to N\Big(\mathbf{0}, Avar(\hat{\theta_u}^h(\hat{W}))\Big) \tag{B.13}$$

where

$$Avar(\hat{\theta_u}^h(\hat{W})) = \{M(u)'WM(u)\}^{-1}M(u)'W\Sigma(u)WM(u)\{M(u)'WM(u)\}^{-1}$$

and $\Sigma(u) = \begin{bmatrix}v_0\Omega(u) & v_1\Omega(u)\\ v_1\Omega(u) & v_2\Omega(u)\end{bmatrix}$; $\Omega(u) = \Gamma_0^h(u) + 2\sum_{s=1}^{\infty}\Gamma_s^h(u)$; *and* $v_s = \int u^s k^2(u)du$.

B.3.3 Asymptotic Variance Estimator

Suppose there is available a estimator for $\Omega(u)$, say $\hat{\Omega}(u)$. A consistent estimator for $Avar(\theta_u^h(\hat{W}))$ is:

$$\hat{Avar}(\theta u(\hat{W})) = \{\hat{M}(u)'\hat{W}M(u)\}^{-1}M(u)'\hat{W}\hat{\Sigma}(u)\hat{W}M(u)\{M(u)'\hat{W}M(u)\}^{-1} \tag{B.14}$$

where $\hat{\Sigma}(u) = \begin{bmatrix} v_0\hat{\Omega}(u) & v_1\hat{\Omega}(u) \\ v_1\hat{\Omega}(u) & v_2\hat{\Omega}(u) \end{bmatrix}$ an the estimator $\hat{M}(u)$ is obtained by replacing $\hat{M}_{zx}(u)$ in: $\hat{M}(u) = \begin{bmatrix} \mu_0\hat{M}_{zx}(u) & \mu_1\hat{M}_{zx}(u) \\ \mu_1\hat{M}_{zx}(u) & \mu_2\hat{M}_{zx}(u) \end{bmatrix}$ where $\hat{M}_{zx}(u) = \frac{1}{n}\sum_{t=0}^{n} k_b(t/n - u)\mathbf{z_t}\mathbf{x_t'}$. There is left to find potential candidates for $\hat{\Omega}$. Cai (2007) proposes:

$$\hat{\Omega}(u) = v_0^{-1}\frac{b}{n}(\sum_{t=1}^{n} k_b(t/n-u)\hat{\epsilon}_{t+h}^h z_t)(\sum_{t=1}^{n} k_b(t/n-u)\hat{\epsilon}_{t+h}^h z_t)' \tag{B.15}$$

where $\hat{\epsilon}_{t+h}^h = y_{t+h} - x_t'\hat{\delta}^h(u)$.

However, this estimator does not have good properties en small samples. Therefore, let's consider an alternative estimator for $\Omega(u)$: a HAC estimator which takes into account that $z_t\hat{\epsilon}_{t+h}^h$ might be serially correlated or/and heteroskedastic. Following the Newey and West (1986) procedure, the HAC estimator for $\Omega(u)$ is:

$$\hat{\Omega}(u) = v_0^{-1}(\hat{J}_0 + \sum_{i=1}^{\lambda}(1 - \frac{i}{\lambda+1})(\hat{J}_i + \hat{J}_{-i})) \tag{B.16}$$

where λ is the bandwidth parameter for the Bartlett kernel and:

$$\hat{J}_s = \frac{b}{n}\sum_{t=s+1}^{n} k_{ut}k_{u(t-s)}x_t x_{t-s}'\hat{\epsilon}_t\hat{\epsilon}_{t-s} \tag{B.17}$$

B.3.4 Proof of Theorem 1

Let's define the residual of a second order Taylor approximation:

$$M(t/n) = \delta^h(t/n) - \delta^h(u) - \delta^{h\prime}(u)(t/n-u) - \frac{1}{2}\delta^{h\prime\prime}(u)(t/n-u)^2 \tag{B.18}$$

From the expression for the estimator in equation B.8:

$$\hat{\theta_u}^h(\hat{W}) - \theta_u^h = (S_{ZX}^{u\prime}\hat{W}S_{ZX}^u)^{-1}S_{ZX}^{u\prime}\hat{W}S_{Zy}^{u,h} - \theta_u^h \tag{B.19}$$

$$\hat{\theta_u}^h(\hat{W}) - \theta_u^h = (S_{ZX}^{u\prime}\hat{W}S_{ZX}^u)^{-1}(S_{ZX}^{u\prime}\hat{W}S_{Zy}^{u,h} - S_{ZX}^{u\prime}\hat{W}S_{ZX}^u\theta_u^h) \tag{B.20}$$

Substituting y_{t+h} and θ_u^h by $x_t'\delta^h(t/n) + \epsilon_{t+h}^h$ and $[\delta^h(u)' \quad \delta^{h\prime}(u)']'$, respectively:

$$\hat{\theta_u}^h(\hat{W}) - \theta_u^h = (S_{ZX}^{u\prime}\hat{W}S_{ZX}^u)^{-1}(S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)z_t(x_t'\delta^h(t/n) + \epsilon_{t+h}^h) -$$

$$S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'\delta_u^h - S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'(t/n-u)\delta_u^{h\prime}) \tag{B.21}$$

Substituting $\delta^h(t/n)$ by equation B.18:

$$\begin{aligned}\hat{\theta_u}^h(\hat{W}) - \theta_u^h = (S_{ZX}^{u\prime}\hat{W}S_{ZX}^u)^{-1}(S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}(x_t'M(t)+ \\ x_t'\delta^h(u) + x_t'\delta^{h\prime}(u)(t/n-u) + \frac{1}{2}x_t'\delta^{h\prime\prime}(u)(t/n-u)^2 + \epsilon_{t+h}^h) - \\ S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'\delta_u^h - \\ S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'(t/n-u)\delta_u^{h\prime})\end{aligned} \tag{B.22}$$

After some terms cancel out we get:

$$\begin{aligned}\hat{\theta_u}^h(\hat{W}) - \theta_u^h = (S_{ZX}^{u\prime}\hat{W}S_{ZX}^u)^{-1}(S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'M(t/n)+ \\ \frac{1}{2}S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}x_t'\delta^{h\prime\prime}(u)(t/n-u)^2+ \\ S_{ZX}^{u\prime}\hat{W}\frac{1}{n}\sum_{t=1}^{n}k_b(t/n-u)Z_{t,u}\epsilon_{t+h}^h)\end{aligned} \tag{B.23}$$

Let's define:

$$
\begin{aligned}
S &= S_{ZX}^{u\prime}\hat{W}S_{ZX}^{u} = S_{ZX}^{u\prime}H_2^{-1}H_2\hat{W}H_2H_2^{-1}S_{ZX}^{u} \\
R &= \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)Z_{t,u}x_t'M(t/n) \\
B &= \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)Z_{t,u}x_t'\delta^{h\prime\prime}(u)(t/n-u)^2 \\
\mathbb{T} &= \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)Z_{t,u}\epsilon_{t+h}^{h}
\end{aligned}
\tag{B.24}
$$

Then, equation B.23 can be rewritten as:

$$
\hat{\theta_u}^{h}(\hat{W}) - \theta_u^h = S^{-1}S_{ZX}^{u\prime}\hat{W}R + S^{-1}S_{ZX}^{u\prime}\hat{W}B + S^{-1}S_{ZX}^{u\prime}\hat{W}\mathbb{T} \tag{B.25}
$$

Scaling by H_1 and defining

$$
\begin{aligned}
H_1(\hat{\theta_u}^{h}(\hat{W}) - \theta_u^h) = (H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}H_2\hat{W}H_2H_2^{-1}R+ \\
(H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}H_2\hat{W}H_2H_2^{-1}B+ \\
(H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}H_2\hat{W}H_2H_2^{-1}\mathbb{T}
\end{aligned}
\tag{B.26}
$$

where $H_1 = diag(1,b)\otimes I_{(d\times d)}$ and $H_2 = diag(1,b)\otimes I_{(m\times m)}$.

Recall that $\hat{W} = H_2^{-1}\hat{W}H_2^{-1}$:

$$
\begin{aligned}
H_1(\hat{\theta_u}^{h}(\hat{W}) - \theta_u^h) = (H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}\hat{W}H_2^{-1}R+ \\
(H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}\hat{W}H_2^{-1}B+ \\
(H_1^{-1}SH_1^{-1})^{-1}H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}\hat{W}H_2^{-1}\mathbb{T}
\end{aligned}
\tag{B.27}
$$

The expression $H_1^{-1}SH_1^{-1}$ can be written as:

$$
\begin{aligned}
H_1^{-1}SH_1^{-1} &= H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}H_2\hat{W}H_2H_2^{-1}S_{ZX}^{u}H_1^{-1} \\
&= H_1^{-1}S_{ZX}^{u\prime}H_2^{-1}\hat{W}H_2^{-1}S_{ZX}^{u}H_1^{-1}
\end{aligned}
\tag{B.28}
$$

Let's define $S(u) = H_2^{-1}S_{ZX}^{u}H_1^{-1}$, therefore equation B.27 can be rewritten as:

$$
\begin{aligned}
H_1(\hat{\theta_u}^{h}(\hat{W}) - \theta_u^h) = \{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}R+ \\
\{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}B + \{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}\mathbb{T}
\end{aligned}
\tag{B.29}
$$

Notice that $\mathbf{S}(\mathbf{u})$ can be written as:

$$\begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t' & \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t' \frac{(t/n-u)}{b} \\ \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t' \frac{(t/n-u)}{b} & \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t' \frac{(t/n-u)^2}{b^2} \end{bmatrix} = \begin{bmatrix} S(u,0) & S(u,1) \\ S(u,1) & S(u,2) \end{bmatrix} \tag{B.30}$$

such that $S(u,j) = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t'(\frac{t/n-u}{b})^j$.

Lemma 1. *Under assumptions 1 to 4:*

$$S(u,j) - \mu_j M_{zx}(u) \to o_p(1) \tag{B.31}$$

where $\mu_s = \int u^s k(u)du$

Proof. Taking expectations:

$$\begin{aligned} \mathbb{E}(S(u,j)) = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)\mathbb{E}(z_t x_t')(\frac{t/n-u}{b})^j = \\ \sum_{t=1}^{n} \frac{k(t/n-u)}{nb} M_{zx}(t/n)(\frac{t/n-u}{b})^j \end{aligned} \tag{B.32}$$

Under assumptions 2 and 3, by the Riemann sum approximation of an integral and a change of variables:

$$\mathbb{E}(S(u,j)) = \int_{-u/b}^{(1-u)/b} u^j k(u) M_{zx}(ub+u)du + o_p(1) \tag{B.33}$$

As $ub \to 0$:

$$\begin{aligned} \mathbb{E}(S(u,j)) = M_{zx}(u)\int_{-u/b}^{(1-u)/b} u^j k(u)du + o_p(1) \\ = \mu_j M_{zx}(u) + o_p(1) \end{aligned} \tag{B.34}$$

Now, let η_{pq} be the $(p,q)th$ element of $S(u,j)$. That is, $\eta_{pq} = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n - u)z_{pt}x_{qt}(\frac{t/n-u}{b})^j$. Then,

$$Var(\eta_{pq}) = n^{-2}\sum_{t=1}^{n} k_b^2(t/n-u)Var(z_{pt}x_{qt})(\frac{t/n-u}{b})^{2j} + 2n^{-2}\sum_{1<i<l}^{n} Cov(z_{pi}x_{qi}, z_{pl}x_{ql})k_b(i/n-u)(\frac{i/n-u}{b})^j \times k_b(l/n-u)(\frac{l/n-u}{b})^j \tag{B.35}$$

$$Var(\eta_{pq}) = I_1 + I_2 \tag{B.36}$$

$$I_1 = n^{-2}\sum_{t=1}^{n} \frac{k^2(\frac{t/n-u}{b})}{b^2} Var(z_{pt}x_{qt})(\frac{t/n-u}{b})^{2j} \tag{B.37}$$

$$I_1 = (nb)^{-1}\sum_{t=1}^{n} \frac{k^2(\frac{t/n-u}{b})}{nb} Var(z_{pt}x_{qt})(\frac{t/n-u}{b})^{2j} \tag{B.38}$$

Let $C = max_t(Var(z_{pt}x_{qt}))$, then by the Riemann sum approximation of an integral:

$$I_1 \leq C(nb)^{-1}\sum_{t=1}^{n} \frac{k^2(\frac{t/n-u}{b})}{nb}(\frac{t/n-u}{b})^{2j} \tag{B.39}$$

$$I_1 \leq C(nb)^{-1}\int_{-u/b}^{(1-u)/b} u^{2j}k^2(u)du \leq C^*(nb)^{-1} = o_p(1) \tag{B.40}$$

Now, let's analyze I_2:

$$I_2 = 2n^{-2}\sum_{1<i<l}^{n} Cov(z_{pi}x_{qi}, z_{pl}x_{ql})k_b(i/n-u)(\frac{i/n-u}{b})^j \times k_b(l/n-u)(\frac{l/n-u}{b})^j \tag{B.41}$$

$$I_2 \leq \mid I_2 \mid \leq 2n^{-2}\sum_{1<i<l}^{n} \mid Cov(z_{pi}x_{qi}, z_{pl}x_{ql}) \mid\mid k_b(i/n-u)(\frac{i/n-u}{b})^j \times \mid\mid k_b(l/n-u)(\frac{l/n-u}{b})^j \mid \tag{B.42}$$

By applying Corollary A.2 in Hall and Heyde (1980) which is cited by Cai (2007). We can find a similar result in Proposition 2.5 in Fan and Yao (2003). Under assumption 2, we have:

$$\begin{aligned} \mid Cov(z_{pi}x_{qi}, z_{pl}x_{ql}) \mid \leq 8\{E \mid z_{pi}x_{qi} \mid^{2+\delta}\}^{\frac{1}{2+\delta}}\{E \mid z_{pl}x_{ql} \mid^{2+\delta}\}^{\frac{1}{2+\delta}} \\ \times\alpha(l-i)^{1-\frac{1}{2+\delta}-\frac{1}{2+\delta}} \end{aligned} \tag{B.43}$$

Moreover, under assumption 4 we get that $E \mid z_{pr}x_{qr} \mid^{2+\delta} < \infty$ for $r = i, l$:

$$\mid Cov(z_{pi}x_{qi}, z_{pl}x_{ql}) \mid \leq C\alpha(l-i)^{\frac{\delta}{2+\delta}} \tag{B.44}$$

Therefore,

$$\begin{aligned} \mid I_2 \mid \leq 2Cn^{-2}\sum_{1<i<l}^{n}\alpha(l-i)^{\frac{\delta}{2+\delta}} \mid k_b(i/n-u)(\frac{i/n-u}{b})^j \\ \times \mid\mid k_b(l/n-u)(\frac{l/n-u}{b})^j \mid \end{aligned} \tag{B.45}$$

Changing the index in the summation:

$$\begin{aligned} \mid I_2 \mid \leq 2Cn^{-2}\sum_{s=1}^{n-1}\alpha(s)^{\frac{\delta}{2+\delta}}\sum_{i=1}^{n-p} \mid k_b(i/n-u)(\frac{i/n-u}{b})^j \\ \times \mid\mid k_b(i/n-u+s/n)(\frac{l/n-u}{b}+\frac{s}{nb})^j \mid \end{aligned} \tag{B.46}$$

$$\begin{aligned} \mid I_2 \mid \leq 2Cn^{-2}\sum_{s=1}^{n-1}\alpha(s)^{\frac{\delta}{2+\delta}}\sum_{i=1}^{n-p} \mid \frac{k(\frac{i/n-u}{b})}{b}(\frac{i/n-u}{b})^j \\ \times \mid\mid \frac{k(\frac{i/n-u}{b}+\frac{s}{nb})}{b}(\frac{l/n-u}{b}+\frac{s}{nb})^j \mid \end{aligned} \tag{B.47}$$

As $n \to \infty$, s/nb goes to zero:

$$\begin{aligned} \mid I_2 \mid \leq 2Cn^{-2}\sum_{s=1}^{n-1}\alpha(s)^{\frac{\delta}{2+\delta}}\sum_{i=1}^{n-p} \mid \frac{k(\frac{i/n-u}{b})}{b}(\frac{i/n-u}{b})^j \\ \times \mid\mid \frac{k(\frac{i/n-u}{b})}{b}(\frac{l/n-u}{b})^j \mid \end{aligned} \tag{B.48}$$

By the Riemann sum approximation of an integral:

$$\mid I_2 \mid \leq 2C(nb)^{-1} \sum_{s=1}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \sum_{i=1}^{n-p} \frac{k^2(\frac{i/n-u}{b})}{nb} (\frac{i/n-u}{b})^{2j} \tag{B.49}$$

$$\mid I_2 \mid \leq 2C(nb)^{-1} \sum_{s=1}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \int_{-u/b}^{(1-u)/b} u^{2j} k^2(u) du \tag{B.50}$$

$$\mid I_2 \mid \leq C^*(nb)^{-1} \sum_{s=1}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \tag{B.51}$$

Given that $\alpha(s)^{\frac{\delta}{2+\delta}} = O(s^{-(1+\delta)})$ by assumption 2, $\sum_{s=1}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} = O(1)$. Then,

$$\mid I_2 \mid \leq C^*(nb)^{-1} \sum_{s=1}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \leq C^{**}(nb)^{-1} = o_p(1) \tag{B.52}$$

Putting the pieces together:

$$\mathbb{E}(S(u,j)) = \mu_j M_{zx}(u) + o_p(1) \tag{B.53}$$

$$Var(S(u,j)) = o_p(1) + o_p(1) = o_p(1) \tag{B.54}$$

Thus,

$$S(u,j) - \mu_j M_{zx}(u) \to o_p(1) \tag{B.55}$$

□

Lemma 2. *Under assumptions 1 to 4:*

$$H_2^{-1} R \to o_p(b^2) \tag{B.56}$$

Proof. Let's consider $b^{-2} H_2^{-1} R$:

$$b^{-2} H_2^{-1} R = \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} k_b(t/n-u) \mathbf{z_t x_t'} b^{-2} M(t/n) \\ \frac{1}{n} \sum_{t=1}^{n} k_b(t/n-u) z_t x_t' \frac{(t/n-u)}{b} b^{-2} M(t/n) \end{bmatrix} = \begin{bmatrix} R_0 \\ R_1 \end{bmatrix} \tag{B.57}$$

such that $R_j = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t'(\frac{t/n-u}{b})^j b^{-2}M(t/n)$. Taking the expectation of last expression we get:

$$\mathbb{E}(R_j) = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)\mathbb{E}(z_t x_t')(\frac{t/n-u}{b})^j b^{-2}M(t/n) \tag{B.58}$$

$$\mathbb{E}(R_j) = \sum_{t=1}^{n} \frac{k(\frac{t/n-u}{b})}{nb}(\frac{t/n-u}{b})^j M_{zx}(t/n)b^{-2}M(t/n) \tag{B.59}$$

By the Riemann sum approximation of an integral:

$$\mathbb{E}(R_j) = \int_{-u/b}^{(1-u)/b} u^j k(u) M_{zx}(ub+u)b^{-2}M(ub+u)du + o_p(1) \tag{B.60}$$

By the Taylor approximation, we know that: $M(ub+u) = o_p(b^2)$ for any u. Therefore $b^{-2}M(ub+u) \to o_p(1)$.

$$\mathbb{E}(R_j) = \mu_j M_{zx}(u)b^{-2}M(ub+u)du + o_p(1) = o_p(1) \tag{B.61}$$

Using the same techniques in the proof of Lemma 1, we have that:

$$Var(R_j) = o_p(1) \tag{B.62}$$

Thus,

$$b^{-2}H_2^{-1}R = o_p(1) \quad then \quad H_2^{-1}R \to o_p(b^2) \tag{B.63}$$

□

Lemma 3.

$$H_2^{-1}B \to \frac{1}{2}b^2 \begin{bmatrix} \mu_2 M_{zx}(u)\delta^{h\prime\prime}(u) \\ \mu_3 M_{zx}(u)\delta^{h\prime\prime}(u) \end{bmatrix} + o_p(b^2) \tag{B.64}$$

Proof.

$$H_2^{-1}B = \frac{1}{2}b^2 \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t'\delta^{h\prime\prime}(u)(\frac{t/n-u}{b})^2 \\ \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t'\delta^{h\prime\prime}(u)(\frac{t/n-u}{b})^3 \end{bmatrix} = \frac{1}{2}b^2 \begin{bmatrix} B_2 \\ B_3 \end{bmatrix} \tag{B.65}$$

Such that $B_j = \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t x_t'\delta^{h\prime\prime}(u)(\frac{t/n-u}{b})^j = S(u,j)\delta''(u)$

From Lemma 1:

$$B_j - \mu_j M_{zx}(u)\delta^{h\prime\prime}(u) = o_p(1) \tag{B.66}$$

Therefore:

$$H_2^{-1}B = \frac{1}{2}b^2 \begin{bmatrix} \mu_2 M_{zx}(u)\delta^{h\prime\prime}(u) \\ \mu_3 M_{zx}(u)\delta^{h\prime\prime}(u) \end{bmatrix} + o_p(b^2) \tag{B.67}$$

□

Lemma 4.

$$(nb)Var(H_2^{-1}\mathbb{T}) - \Sigma(u) \to o_p(1) \tag{B.68}$$

where $v_s = \int u^s k^2(u)du$; $\Omega(u) = \Gamma_0^h(u) + 2\sum_{s=1}^{\infty}\Gamma_s^h(u)$; *and* $\Sigma(u) = \begin{bmatrix} v_0 & v_1 \\ v_1 & v_2 \end{bmatrix} \otimes \Omega(u)$

Proof.

$$H_2^{-1}\mathbb{T} = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t\epsilon_{t+h}^h \\ \frac{1}{n}\sum_{t=1}^{n} k_b(t/n-u)z_t\epsilon_{t+h}^h(\frac{t/n-u}{b}) \end{bmatrix} = \begin{bmatrix} \mathbb{T}_0 \\ \mathbb{T}_1 \end{bmatrix} \tag{B.69}$$

$$(nb)Var(H_2^{-1}\mathbb{T}) = \begin{bmatrix} (nb)Var(\mathbb{T}_0) & (nb)Cov(\mathbb{T}_0,\mathbb{T}_1) \\ (nb)Cov(\mathbb{T}_0,\mathbb{T}_1)' & (nb)Var(\mathbb{T}_1) \end{bmatrix} \tag{B.70}$$

Let's define, $\Gamma_k^h(t/n) = Cov(z_t\epsilon_{t+h}^h, z_{t+k}\epsilon_{t+h+k}^h)$ and consider $\mathbb{T}_0$:

$$\begin{aligned} (nb)Var(\mathbb{T}_0) &= n^{-1}b\sum_{t=1}^{n} Cov(z_t\epsilon_{t+h}^h, z_t\epsilon_{t+h}^h)k_b^2(t/n-u) \\ +2n^{-1}b\sum_{1\le l<k\le n} &Cov(z_l\epsilon_{l+h}^h, z_k\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) \\ &= I_3 + I_4 \end{aligned} \tag{B.71}$$

First, consider I_3:

$$\begin{aligned} I_3 = n^{-1}b\sum_{t=1}^{n} Cov(z_t\epsilon_{t+h}^h, z_t\epsilon_{t+h}^h)k_b^2(t/n-u) = \\ \sum_{t=1}^{n}\Gamma_0^h(t/n)\frac{k^2(\frac{t/n-u}{b})}{nb} \end{aligned} \tag{B.72}$$

Under assumption 3, by the Riemann sum approximation of an integral and taking into account $ub \to 0$:

$$I_3 = \int_{-u/b}^{(1-u)/b} k^2(u)\Gamma_0^h(ub+u)du + o_p(1) = \Gamma_0^h(u)\int_{-u/b}^{(1-u)/b} k^2(u)du + o_p(1) = \Gamma_0^h(u)v_0 + o_p(1) \tag{B.73}$$

For I_4 we follow Cai (2007) and Chen (2015). There exists $d_T \to \infty$ such that $d_T/\sqrt{n} \to 0$ and $d_T/(nb) \to 0$:

$$I_4 = 2n^{-1}b \sum_{1\le l<k\le n} Cov(\mathbf{z_l}\epsilon_{l+h}^h, \mathbf{z_k}\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) \tag{B.74}$$

$$I_4 = 2n^{-1}b \sum_{l=1}^{n-1}\sum_{k=l+1}^{n} Cov(\mathbf{z_l}\epsilon_{l+h}^h, \mathbf{z_k}\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) \tag{B.75}$$

$$\begin{aligned} I_4 = {} & 2n^{-1}b \sum_{l=1}^{n-1}\sum_{1\le k-l\le d_T} Cov(\mathbf{z_l}\epsilon_{l+h}^h, \mathbf{z_k}\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) + \\ & 2n^{-1}b \sum_{l=1}^{n-1}\sum_{d_T\le k-l\le n-1} Cov(\mathbf{z_l}\epsilon_{l+h}^h, \mathbf{z_k}\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) \\ & = I_{41} + I_{42} \end{aligned} \tag{B.76}$$

Let's work out I_{42}:

$$I_{42} = 2n^{-1}b \sum_{l=1}^{n-1}\sum_{d_T\le k-l\le n-1} Cov(\mathbf{z_l}\epsilon_{l+h}^h, \mathbf{z_k}\epsilon_{k+h}^h)k_b(l/n-u)k_b(k/n-u) \tag{B.77}$$

For any $(p,q)th$ element of $I_{42_{(p,q)}}$, we have:

$$\begin{aligned} | I_{42_{(p,q)}} | = {} & 2n^{-1}\sum_{l=1}^{n-1} | k_b(l/n-u) | \\ & \times \sum_{d_T\le k-l\le n-1} | Cov(z_{pl}\epsilon_{l+h}^h, z_{qk}\epsilon_{k+h}^h) | \, | bk_b(k/n-u) | \end{aligned} \tag{B.78}$$

$$\begin{aligned} \mid I_{42_{(p,q)}} \mid &= 2\sum_{l=1}^{n-1} \mid \frac{k(\frac{l/n-u}{b})}{nb} \\ \times \mid \sum_{d_T \leq k-l \leq n-1} &\mid Cov(z_{pl}\epsilon_{l+h}^h, z_{qk}\epsilon_{k+h}^h) \mid\mid k(\frac{k/n-u}{b}) \mid \end{aligned} \tag{B.79}$$

Under assumption 2, by applying Corollary A.2 in Hall and Heyde (1980) which is cited by Cai (2007). We can find a similar result in Proposition 2.5 in Fan and Yao (2003).

$$\begin{aligned} \mid Cov(z_{pl}\epsilon_{l+h}^h, z_{qk}\epsilon_{k+h}^h) \mid &\leq 8\{E \mid z_{pl}\epsilon_{l+h}^h \mid^{2+\delta}\}^{\frac{1}{2+\delta}} \\ &\times\{E \mid z_{qk}\epsilon_{k+h}^h \mid^{2+\delta}\}^{\frac{1}{2+\delta}} \alpha(k-l)^{1-\frac{1}{2+\delta}-\frac{1}{2+\delta}} \end{aligned} \tag{B.80}$$

Letting C be a constant term:

$$\mid Cov(z_{pl}\epsilon_{l+h}^h, z_{qk}\epsilon_{k+h}^h) \mid \leq C\alpha(k-l)^{\frac{\delta}{2+\delta}} \tag{B.81}$$

Moreover, we know that $k(.)$ is bounded by assumption 4. By defining $C^* = max_k\{k(\frac{k/n-u}{b})\}$, we can guarantee:

$$\mid I_{42_{(p,q)}} \mid \leq 2CC^* \sum_{l=1}^{n-1} \mid \frac{k(\frac{l/n-u}{b})}{nb} \mid \sum_{d_T \leq k-l \leq n-1} \alpha(k-l)^{\frac{\delta}{2+\delta}} \tag{B.82}$$

$$\mid I_{42_{(p,q)}} \mid \leq 2CC^* \sum_{l=1}^{n-1} \mid \frac{k(\frac{l/n-u}{b})}{nb} \mid \sum_{s=d_T}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \tag{B.83}$$

By the Riemann sum approximation of an integral and considering that $\alpha(s) = O(s^{-\frac{(2+\delta)(1+\delta)}{\delta}})$ by assumption 2. Notice that $\sum_{s=d_T}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} = \sum_{s=d_T}^{n-1} O(s^{-(1+\delta)}) = O(d_T^{-(1+\delta)})$.

$$\mid I_{42_{(p,q)}} \mid \leq 2CC^* \int_{-u/b}^{(1-u)/b} k(u)du \sum_{s=d_T}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \tag{B.84}$$

$$\mid I_{42_{(p,q)}} \mid \leq 2CC^*\mu_0 \sum_{s=d_T}^{n-1} \alpha(s)^{\frac{\delta}{2+\delta}} \leq C^{**} d_T^{-(1+\delta)} \leq C^{**} d_T^{-\delta} = o_p(1) \tag{B.85}$$

which follows from $d_T \to \infty$.

Now consider $(p,q)th$ element of $I41_{(p,q)}$, we have:

$$I41_{(p,q)} = 2n^{-1}b\sum_{l=1}^{n-1}\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \times k_b(l/n-u)k_b(k/n-u) \tag{B.86}$$

$$I41_{(p,q)} = 2n^{-1}b\sum_{l=1}^{n-1}\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \times(k_b(l/n-u)k_b(k/n-u) - k_b^2(l/n-u) + k_b^2(l/n-u)) \tag{B.87}$$

$$\begin{aligned} I41_{(p,q)} &= 2n^{-1}b\sum_{l=1}^{n-1}\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \\ &\quad\times(k_b(l/n-u)k_b(k/n-u) - k_b^2(l/n-u))+ \\ &\quad 2n^{-1}b\sum_{l=1}^{n-1}\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h})k_b^2(l/n-u) \\ &= I411_{(p,q)} + I412_{(p,q)} \end{aligned} \tag{B.88}$$

Then:

$$I412_{(p,q)} = 2n^{-1}b\sum_{l=1}^{n-1} k_b^2(l/n-u)\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \tag{B.89}$$

By the definition $\Gamma^h_k(t/n) = Cov(z_t\epsilon^h_{t+h}, z_{t+k}\epsilon^h_{t+h+k})$:

$$I412_{(p,q)} = 2\sum_{l=1}^{n-1}\frac{k^2(\frac{l/n-u}{b})}{nb}\sum_{s=1}^{d_T}\Gamma^h_s(l/n)_{(p,q)} \tag{B.90}$$

By the Riemann sum approximation of an integral and noticing that $ub \to 0$ and $d_T \to \infty$ as $n \to \infty$:

$$\begin{aligned} I412_{(p,q)} &= 2\int_{-u/b}^{(1-u)/b} k^2(u)\sum_{s=1}^{d_T}\Gamma^h_s(ub+u)_{(p,q)} + o_p(1) \\ &= 2v_0\sum_{s=1}^{\infty}\Gamma^h_s(u)_{(p,q)} + o_p(1) \end{aligned} \tag{B.91}$$

Now, $I_{411_{(p,q)}}$:

$$I_{411_{(p,q)}} = 2n^{-1}b\sum_{l=1}^{n-1}\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \times(k_b(l/n-u)k_b(k/n-u) - k_b^2(l/n-u)) \tag{B.92}$$

$$I_{411_{(p,q)}} = 2n^{-1}b\sum_{l=1}^{n-1} k_b(l/n-u)\sum_{1\leq k-l\leq d_T} Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \times(k_b(k/n-u) - k_b(l/n-u)) \tag{B.93}$$

$$\mid I_{411_{(p,q)}} \mid \leq 2n^{-1}b\sum_{l=1}^{n-1} k_b(l/n-u) \times \sum_{1\leq k-l\leq d_T} \mid Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \mid\mid k_b(k/n-u) - k_b(l/n-u) \mid \tag{B.94}$$

By assumption 5, $k(.)$ satisfies the Lipschitz condition, thus:

$$\mid k_b(k/n-u) - k_b(l/n-u) \mid \leq C\frac{k-l}{nb^2} \leq C\frac{d_T}{nb^2} \tag{B.95}$$

Then,

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-2}b^{-1}d_T\sum_{l=1}^{n-1} k_b(l/n-u)\sum_{1\leq k-l\leq d_T} \mid Cov(z_{pl}\epsilon^h_{l+h}, z_{qk}\epsilon^h_{k+h}) \mid \tag{B.96}$$

Adding and subtracting a covariance term at period Tu:

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-2}b^{-1}d_T\sum_{l=1}^{n-1} k_b(l/n-u) \times\sum_{s=1}^{d_T} \mid Cov(z_{pl}\epsilon^h_{l+h}, z_{ql+s}\epsilon^h_{l+s+h}) - Cov(z_{pt}\epsilon^h_{t+h}, z_{qt+s}\epsilon^h_{t+s+h}) + Cov(z_{pt}\epsilon^h_{t+h}, z_{qt+s}\epsilon^h_{t+s+h}) \mid \tag{B.97}$$

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-2}b^{-1}d_T \sum_{l=1}^{n-1} k_b(l/n-u)$$
$$\times \sum_{s=1}^{d_T} (\mid Cov(z_{pl}\epsilon^h_{l+h}, z_{ql+s}\epsilon^h_{l+s+h}) - Cov(z_{pt}\epsilon^h_{t+h}, z_{qt+s}\epsilon^h_{t+s+h}) \mid + \tag{B.98}$$
$$\mid Cov(z_{pt}\epsilon^h_{t+h}, z_{qt+s}\epsilon^h_{t+s+h}) \mid)$$

Under $\mid Cov(z_{pl}\epsilon^h_{l+h}, z_{ql+s}\epsilon^h_{l+s+h}) - Cov(z_{pt}\epsilon^h_{t+h}, z_{qt+s}\epsilon^h_{t+s+h}) \mid = O(b)$:

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-2}b^{-1}d_T \sum_{l=1}^{n-1} k_b(l/n-u) \sum_{s=1}^{d_T} (bC^* + \mid \Gamma^h_s(u)_{(j,m)} \mid) \tag{B.99}$$

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-2}b^{-1}d_T \sum_{l=1}^{n-1} k_b(l/n-u)$$
$$\times (\sum_{s=1}^{d_T} \mid \Gamma^h_s(u)_{(j,m)} \mid + d_T bC^*) \tag{B.100}$$

$$\mid I_{411_{(p,q)}} \mid \leq Cn^{-1}b^{-1}d_T \sum_{l=1}^{n-1} \frac{k(\frac{l/n-u}{b})}{nb}$$
$$\times (\sum_{s=1}^{d_T} \mid \Gamma^h_s(u)_{(j,m)} \mid + d_T bC^*) \tag{B.101}$$

$$\mid I_{411_{(p,q)}} \mid \leq \sum_{l=1}^{n-1} \frac{k(\frac{l/n-u}{b})}{nb}$$
$$\times (Cn^{-1}b^{-1}d_T \sum_{s=1}^{d_T} \mid \Gamma^h_s(u)_{(j,m)} \mid + Cn^{-1}d_T^2 C^*) \tag{B.102}$$

We know that as $n \to \infty$, $d_T/\sqrt{n} \to 0$ and $d_T/(nb) \to 0$:

$$
\begin{aligned}
\mid I_{411_{(j,m)}} \mid &\leq \mu_0 (C \sum_{s=1}^{\infty} \mid \Gamma_s^h(u)_{(j,m)} \mid n^{-1} b^{-1} d_T + CC^* n^{-1} d_T^2) \\
&\leq C^{**} (\frac{d_T}{nb} + \frac{d_T^2}{n}) = o_p(1)
\end{aligned}
\tag{B.103}
$$

Putting the pieces together:

$$
(nb) Var(\mathbb{T}_0) = I_3 + I_4 = I_3 + I_{41} + I_{42} = I_3 + I_{411} + I_{412} + I_{42} \tag{B.104}
$$

$$
(nb) Var(\mathbb{T}_0) = v_0 (\Gamma_0^h(u) + 2 \sum_{s=1}^{\infty} \Gamma_s^h(u)) + o_p(1) = v_0 \Omega(u) + o_p(1) \tag{B.105}
$$

Analogously, we have:

$$
(nb) Var(\mathbb{T}_1) = v_2 (\Gamma_0^h(u) + 2 \sum_{s=1}^{\infty} \Gamma_s^h(u)) + o_p(1) = v_2 \Omega(u) + o_p(1) \tag{B.106}
$$

$$
(nb) Cov(\mathbb{T}_0, \mathbb{T}_1) = v_1 (\Gamma_0^h(u) + 2 \sum_{s=1}^{\infty} \Gamma_s^h(u)) + o_p(1) = v_1 \Omega(u) + o_p(1) \tag{B.107}
$$

Therefore,

$$
(nb) Var(H_2^{-1} \mathbb{T}) - \begin{bmatrix} v_0 & v_1 \\ v_1 & v_2 \end{bmatrix} \otimes \Omega(u) \to o_p(1) \tag{B.108}
$$

by Chebyshev's inequality the result above implies:

$$
H_2^{-1} \mathbb{T} \to O_p((nb)^{-1/2}) \tag{B.109}
$$

□

Under Lemma 1:

$$
S(u) = \begin{bmatrix} \mu_0 M_{zx}(u) & \mu_1 M_{zx}(u) \\ \mu_1 M_{zx}(u) & \mu_2 M_{zx}(u) \end{bmatrix} + o_p(1) = M(u) + o_p(1) = O_p(1) \tag{B.110}
$$

We also know that $\hat{W} - W = o_p(1)$. Therefore, under lemmas 1-4 the equation B.29:

$$\begin{aligned} H_1(\hat{\theta_u}^h(\hat{W}) - \theta_u^h) = \{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}R+ \\ \{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}B+ \\ \{S(u)'\hat{W}S(u)\}^{-1}S(u)'\hat{W}H_2^{-1}\mathbb{T} \end{aligned} \tag{B.111}$$

It converges to:

$$\begin{aligned} H_1(\hat{\theta_u}^h(\hat{W}) - \theta_u^h) = o_p(b^2) + \{M(u)'WM(u)\}^{-1}M(u)'WH_2^{-1}B \\ +O_p((nb)^{-1/2}) \end{aligned} \tag{B.112}$$

B.3.5 Proof of Theorem 2

To show Theorem 2 is sufficient to prove that:

$$\sqrt{nb}H_2^{-1}\mathbb{T} \to N(0, \Sigma(u)) \tag{B.113}$$

To prove this we use the Cramer-Wold device. It means that we need to show that for any unit vector $\mathbf{d}$ in $\mathbb{R}^{2m}$:

$$\sqrt{nb}\mathbf{d}'H_2^{-1}\mathbb{T} \to N(0, \mathbf{d}'\Sigma(u)\mathbf{d}) \tag{B.114}$$

Let's define $P_t = n^{-1/2}b^{1/2}\mathbf{d}'H_2^{-1}k_b(t/n - u)Z_{t,u}\epsilon_{t+h}^h$. Then, $\sqrt{nb}\mathbf{d}'H_2^{-1}\mathbb{T} = \sum_{t=1}^n P_t$. By Lemma 4:

$$Var(\sum_{t=1}^{n} P_t) = \mathbf{d}'\Sigma(u)\mathbf{d} + o_p(1) \tag{B.115}$$

Here we use the small-block large-block technique used in Cai (2007). That is, let's build a partition of $\{1, ..., n\}$ into $2q_T$ subsets: q_T subsets with large-block of size $r_T = \lfloor (nb)^{1/2} \rfloor$; q_T subsets with small-block of size $s_T = \lfloor (nb)^{1/2}/\log(n) \rfloor$; and one last subset with the remainders. The value of $q_T = \lfloor n/(r_T + s_T) \rfloor$. Take $r_j^* = j(r_T + s_T)$ and define the random variables, for $0 \leq j \leq q_T - 1$:

$$\eta_j = \sum_{t=r_j^*+1}^{r_j^*+r_T} P_t, \quad \xi_j = \sum_{t=r_j^*+r_T+1}^{r_{j+1}^*} P_t \quad and \quad Q_3 = \sum_{t=r_j^*+r_T+1}^{n} P_t \tag{B.116}$$

then, $\sqrt{nb}\mathbf{d}'H_2^{-1}n = Q_1 + Q_2 + Q_3$, where $Q_1 = \sum_{j=0}^{q_T-1}\eta_j$ and $Q_2 = \sum_{j=0}^{q_T-1}\xi_j$. Let's take the following results from Cai (2007):

$$\mathbb{E}(Q_2)^2 \to 0, \quad \mathbb{E}(Q_3)^2 \to 0, \tag{B.117}$$

For any s:

$$\mid \mathbb{E}(e^{isQ_1}) - \prod_{j=0}^{q_T-1} \mathbb{E}(e^{is\eta_j}) \mid \to 0 \tag{B.118}$$

and

$$Var(Q_1) \to \mathbf{d}'\Sigma(u)\mathbf{d} \quad and \quad \sum_{j=0}^{q_T-1} \mathbb{E}(\mid \eta_j \mid^{2+\delta}) \to 0 \tag{B.119}$$

The results in B.117 imply that Q_2 and Q_3 are asymptotically irrelevant. The result in B.118 means that $\eta's$ are asymptotically independent and the result in B.119 refers to the standard Lindeberg-Feller and Lyapounov conditions for asymptotic normality. Cai (2007) shows that under assumptions 1-6, results B.117,B.118, and B.119 are accomplished.

B.4 Tables and Figures

Table B.1: Coverage and Length for 90 Confidence Interval with n=500 in Experiment 1

	Inflation																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.99	1.00	0.85	0.96	1.00	0.88	1.00	1.00	0.86	0.22	0.86	0.24	1.09	1.28	0.27	2.12	2.25	0.33
2	0.98	1.00	0.86	0.75	1.00	0.87	0.79	1.00	0.87	0.29	0.86	0.34	0.52	1.28	0.37	0.88	2.25	0.43
3	0.97	1.00	0.86	0.70	1.00	0.85	0.71	1.00	0.87	0.31	0.86	0.38	0.38	1.28	0.41	0.65	2.25	0.47
4	0.96	1.00	0.87	0.85	1.00	0.85	0.73	1.00	0.88	0.36	0.86	0.44	0.47	1.28	0.46	0.66	2.25	0.52
5	0.94	1.00	0.86	0.90	0.98	0.87	0.97	1.00	0.86	0.41	0.86	0.50	1.15	1.28	0.52	2.11	2.25	0.55
6	0.91	1.00	0.88	0.86	1.00	0.86	0.90	1.00	0.86	0.45	0.86	0.51	0.67	1.28	0.51	1.09	2.25	0.53
7	0.86	1.00	0.89	0.87	1.00	0.86	0.86	1.00	0.87	0.49	0.86	0.52	0.46	1.28	0.51	0.46	2.25	0.53
8	0.87	0.99	0.86	0.83	1.00	0.87	0.83	1.00	0.87	0.54	0.86	0.52	0.53	1.28	0.53	0.62	2.25	0.56
9	0.88	0.98	0.85	0.90	0.99	0.87	0.96	1.00	0.85	0.57	0.86	0.51	0.98	1.28	0.51	1.71	2.25	0.51
10	0.87	0.97	0.85	0.87	0.99	0.86	0.91	1.00	0.84	0.59	0.86	0.52	0.81	1.28	0.54	1.27	2.25	0.56
11	0.86	0.96	0.84	0.92	0.97	0.85	0.95	1.00	0.85	0.60	0.86	0.57	0.97	1.28	0.60	1.63	2.25	0.63
12	0.86	0.94	0.85	0.87	0.99	0.86	0.84	1.00	0.84	0.63	0.86	0.65	0.66	1.28	0.67	0.70	2.25	0.69
13	0.85	0.92	0.84	0.87	0.99	0.85	0.86	1.00	0.84	0.69	0.86	0.69	0.72	1.28	0.71	0.74	2.25	0.74
14	0.85	0.89	0.83	0.85	0.99	0.86	0.83	1.00	0.83	0.75	0.86	0.73	0.78	1.28	0.75	0.85	2.25	0.78
15	0.85	0.87	0.84	0.91	0.97	0.85	0.91	1.00	0.82	0.79	0.86	0.76	0.94	1.28	0.78	1.27	2.25	0.80
16	0.85	0.87	0.84	0.85	0.97	0.84	0.82	1.00	0.84	0.82	0.86	0.78	0.88	1.28	0.80	0.98	2.25	0.82
17	0.85	0.86	0.84	0.86	0.97	0.85	0.88	1.00	0.84	0.82	0.86	0.77	0.85	1.28	0.79	0.90	2.25	0.81
18	0.83	0.87	0.84	0.84	0.98	0.83	0.86	1.00	0.85	0.80	0.86	0.78	0.82	1.28	0.80	0.85	2.25	0.81
19	0.82	0.89	0.83	0.81	0.98	0.84	0.82	1.00	0.84	0.76	0.86	0.79	0.77	1.28	0.80	0.80	2.25	0.82
20	0.83	0.92	0.82	0.77	0.99	0.83	0.75	1.00	0.84	0.71	0.86	0.80	0.75	1.28	0.81	0.91	2.25	0.83

Table B.2: Coverage and Length for 90 Confidence Interval with n=500 in Experiment 1

	GDP																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.93	1.00	0.88	0.99	1.00	0.88	0.86	1.00	0.86	0.21	0.50	0.18	0.94	1.42	0.19	1.64	2.78	0.22
2	0.91	1.00	0.88	1.00	1.00	0.87	0.91	1.00	0.88	0.20	0.50	0.20	1.30	1.42	0.20	2.44	2.78	0.24
3	0.91	1.00	0.86	0.91	1.00	0.88	0.87	1.00	0.86	0.23	0.50	0.22	0.51	1.42	0.24	0.91	2.78	0.29
4	0.90	0.99	0.86	0.90	1.00	0.87	0.87	1.00	0.87	0.25	0.50	0.23	0.41	1.42	0.25	0.91	2.78	0.31
5	0.89	0.99	0.86	0.89	1.00	0.85	0.94	1.00	0.86	0.27	0.50	0.25	0.45	1.42	0.28	1.12	2.78	0.34
6	0.88	0.99	0.85	0.91	1.00	0.85	0.86	1.00	0.85	0.28	0.50	0.26	0.33	1.42	0.29	0.54	2.78	0.35
7	0.88	1.00	0.85	0.91	1.00	0.86	0.87	1.00	0.88	0.26	0.50	0.26	0.30	1.42	0.27	0.42	2.78	0.31
8	0.96	1.00	0.85	0.69	1.00	0.85	0.81	1.00	0.86	0.24	0.50	0.26	0.35	1.42	0.27	0.73	2.78	0.31
9	0.99	1.00	0.86	0.66	1.00	0.88	0.76	1.00	0.91	0.23	0.50	0.25	0.27	1.42	0.27	0.55	2.78	0.29
10	1.00	1.00	0.89	0.78	1.00	0.88	0.72	1.00	0.88	0.22	0.50	0.26	0.28	1.42	0.27	0.40	2.78	0.30
11	1.00	1.00	0.88	0.93	1.00	0.87	0.92	1.00	0.87	0.22	0.50	0.26	0.95	1.42	0.27	1.73	2.78	0.31
12	1.00	1.00	0.87	0.96	1.00	0.88	0.92	1.00	0.86	0.23	0.50	0.26	1.24	1.42	0.28	2.27	2.78	0.32
13	1.00	1.00	0.88	0.85	1.00	0.88	0.86	1.00	0.86	0.22	0.50	0.26	0.77	1.42	0.29	1.40	2.78	0.35
14	1.00	1.00	0.87	0.69	1.00	0.87	0.74	1.00	0.87	0.23	0.50	0.26	0.34	1.42	0.29	0.60	2.78	0.37
15	1.00	1.00	0.88	0.70	1.00	0.87	0.74	1.00	0.88	0.22	0.50	0.26	0.34	1.42	0.29	0.62	2.78	0.36
16	1.00	1.00	0.86	0.89	1.00	0.86	0.93	1.00	0.89	0.23	0.50	0.26	0.93	1.42	0.28	1.79	2.78	0.35
17	1.00	1.00	0.87	0.78	1.00	0.86	0.87	1.00	0.87	0.22	0.50	0.26	0.65	1.42	0.29	1.24	2.78	0.35
18	1.00	1.00	0.87	0.76	1.00	0.88	0.82	1.00	0.87	0.22	0.50	0.26	0.55	1.42	0.29	1.01	2.78	0.36
19	1.00	1.00	0.87	0.70	1.00	0.90	0.69	1.00	0.88	0.23	0.50	0.26	0.31	1.42	0.29	0.51	2.78	0.36
20	1.00	1.00	0.88	0.72	1.00	0.86	0.75	1.00	0.90	0.22	0.50	0.26	0.34	1.42	0.29	0.59	2.78	0.36

Table B.3: Coverage and Length for 90 Confidence Interval with n=250 in Experiment 2

	Inflation																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.94	1.00	0.85	0.90	1.00	0.87	0.88	1.00	0.84	0.28	0.95	0.25	0.31	0.98	0.26	0.39	1.13	0.27
2	0.90	1.00	0.84	0.90	1.00	0.83	0.87	1.00	0.85	0.31	0.95	0.30	0.34	0.98	0.31	0.41	1.13	0.31
3	0.89	1.00	0.84	0.90	1.00	0.83	0.88	1.00	0.85	0.34	0.95	0.31	0.37	0.98	0.32	0.44	1.13	0.32
4	0.88	1.00	0.85	0.87	1.00	0.85	0.86	1.00	0.83	0.37	0.95	0.35	0.39	0.98	0.36	0.46	1.13	0.36
5	0.87	0.99	0.83	0.87	1.00	0.82	0.85	1.00	0.84	0.40	0.95	0.41	0.43	0.98	0.41	0.49	1.13	0.41
6	0.85	0.99	0.82	0.89	0.99	0.83	0.85	1.00	0.81	0.44	0.95	0.43	0.47	0.98	0.43	0.53	1.13	0.43
7	0.84	0.98	0.81	0.85	0.99	0.83	0.85	0.99	0.80	0.50	0.95	0.47	0.51	0.98	0.47	0.57	1.13	0.47
8	0.84	0.97	0.80	0.84	0.98	0.81	0.84	0.99	0.80	0.55	0.95	0.52	0.56	0.98	0.52	0.61	1.13	0.52
9	0.83	0.94	0.80	0.84	0.96	0.82	0.83	0.97	0.80	0.59	0.95	0.57	0.60	0.98	0.56	0.65	1.13	0.56
10	0.82	0.93	0.80	0.83	0.95	0.81	0.82	0.97	0.78	0.63	0.95	0.60	0.64	0.98	0.59	0.69	1.13	0.60
11	0.81	0.92	0.80	0.81	0.94	0.80	0.82	0.95	0.79	0.67	0.95	0.63	0.67	0.98	0.62	0.72	1.13	0.63
12	0.80	0.91	0.79	0.81	0.93	0.79	0.80	0.95	0.77	0.70	0.95	0.66	0.71	0.98	0.65	0.76	1.13	0.66
13	0.79	0.91	0.78	0.79	0.92	0.78	0.79	0.96	0.77	0.72	0.95	0.67	0.73	0.98	0.66	0.77	1.13	0.67
14	0.77	0.92	0.78	0.78	0.93	0.78	0.78	0.96	0.77	0.71	0.95	0.67	0.72	0.98	0.66	0.75	1.13	0.67
15	0.77	0.95	0.78	0.77	0.94	0.78	0.78	0.97	0.78	0.68	0.95	0.68	0.69	0.98	0.67	0.72	1.13	0.68
16	0.82	0.97	0.80	0.81	0.97	0.78	0.81	0.99	0.77	0.65	0.95	0.68	0.66	0.98	0.68	0.70	1.13	0.67
17	0.92	0.99	0.80	0.87	0.98	0.77	0.85	0.99	0.78	0.64	0.95	0.68	0.65	0.98	0.68	0.69	1.13	0.67
18	0.96	0.99	0.78	0.90	0.99	0.78	0.89	0.99	0.77	0.65	0.95	0.69	0.64	0.98	0.69	0.71	1.13	0.69
19	0.98	1.00	0.77	0.95	1.00	0.78	0.91	1.00	0.77	0.65	0.95	0.70	0.66	0.98	0.69	0.71	1.13	0.69
20	0.98	1.00	0.79	0.96	1.00	0.79	0.91	1.00	0.77	0.67	0.95	0.70	0.67	0.98	0.70	0.71	1.13	0.70

Table B.4: Coverage and Length for 90 Confidence Interval with n=250 in Experiment 2

	GDP																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.90	1.00	0.83	0.90	1.00	0.84	0.90	1.00	0.85	0.30	0.79	0.24	0.32	0.86	0.24	0.37	1.06	0.26
2	0.91	1.00	0.83	0.89	1.00	0.82	0.88	1.00	0.83	0.30	0.79	0.29	0.32	0.86	0.29	0.39	1.06	0.30
3	0.88	0.99	0.83	0.88	1.00	0.82	0.86	1.00	0.83	0.36	0.79	0.34	0.38	0.86	0.35	0.43	1.06	0.35
4	0.86	0.99	0.85	0.85	0.99	0.81	0.86	0.99	0.82	0.38	0.79	0.38	0.41	0.86	0.38	0.46	1.06	0.38
5	0.84	0.98	0.82	0.82	0.99	0.80	0.83	1.00	0.80	0.41	0.79	0.40	0.43	0.86	0.40	0.48	1.06	0.41
6	0.85	0.98	0.81	0.83	0.98	0.79	0.86	0.99	0.80	0.43	0.79	0.41	0.45	0.86	0.41	0.51	1.06	0.42
7	0.83	0.98	0.81	0.81	0.98	0.79	0.81	0.99	0.78	0.44	0.79	0.42	0.46	0.86	0.41	0.51	1.06	0.42
8	0.81	0.99	0.82	0.81	0.99	0.78	0.81	1.00	0.79	0.43	0.79	0.42	0.45	0.86	0.41	0.50	1.06	0.42
9	0.82	0.99	0.83	0.80	1.00	0.80	0.82	1.00	0.80	0.41	0.79	0.41	0.43	0.86	0.41	0.49	1.06	0.42
10	0.87	1.00	0.81	0.84	1.00	0.83	0.82	1.00	0.82	0.40	0.79	0.42	0.42	0.86	0.41	0.48	1.06	0.42
11	0.92	1.00	0.80	0.89	1.00	0.83	0.87	1.00	0.81	0.40	0.79	0.42	0.43	0.86	0.42	0.50	1.06	0.42
12	0.95	1.00	0.82	0.91	1.00	0.83	0.88	1.00	0.81	0.40	0.79	0.42	0.43	0.86	0.43	0.49	1.06	0.43
13	0.96	1.00	0.80	0.92	1.00	0.82	0.88	1.00	0.81	0.41	0.79	0.42	0.42	0.86	0.42	0.49	1.06	0.43
14	0.98	1.00	0.81	0.93	1.00	0.83	0.87	1.00	0.84	0.41	0.79	0.42	0.42	0.86	0.42	0.48	1.06	0.43
15	0.98	1.00	0.84	0.93	1.00	0.84	0.83	1.00	0.83	0.38	0.79	0.43	0.42	0.86	0.42	0.46	1.06	0.43
16	0.99	1.00	0.84	0.91	1.00	0.85	0.86	1.00	0.84	0.37	0.79	0.42	0.40	0.86	0.43	0.48	1.06	0.43
17	1.00	1.00	0.84	0.92	1.00	0.86	0.82	1.00	0.85	0.38	0.79	0.43	0.41	0.86	0.43	0.47	1.06	0.44
18	1.00	1.00	0.84	0.89	1.00	0.84	0.84	1.00	0.83	0.39	0.79	0.43	0.42	0.86	0.43	0.49	1.06	0.44
19	1.00	1.00	0.85	0.89	1.00	0.84	0.84	1.00	0.83	0.40	0.79	0.43	0.39	0.86	0.43	0.49	1.06	0.44
20	1.00	1.00	0.84	0.89	1.00	0.87	0.81	1.00	0.85	0.39	0.79	0.44	0.39	0.86	0.44	0.48	1.06	0.45

Table B.5: Coverage and Length for 90 Confidence Interval with n=500 in Experiment 2

	Inflation																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.94	1.00	0.86	0.91	1.00	0.86	0.87	1.00	0.84	0.20	0.66	0.19	0.24	0.72	0.19	0.35	0.92	0.20
2	0.93	1.00	0.86	0.89	1.00	0.87	0.89	1.00	0.86	0.23	0.66	0.23	0.26	0.72	0.23	0.36	0.92	0.24
3	0.91	1.00	0.87	0.91	1.00	0.89	0.89	1.00	0.86	0.25	0.66	0.23	0.29	0.72	0.23	0.36	0.92	0.24
4	0.90	1.00	0.87	0.88	1.00	0.86	0.89	1.00	0.86	0.27	0.66	0.26	0.31	0.72	0.26	0.39	0.92	0.27
5	0.89	1.00	0.86	0.87	1.00	0.87	0.86	1.00	0.86	0.30	0.66	0.30	0.33	0.72	0.31	0.41	0.92	0.31
6	0.88	0.99	0.86	0.89	0.99	0.88	0.89	1.00	0.87	0.33	0.66	0.32	0.36	0.72	0.32	0.43	0.92	0.33
7	0.87	0.99	0.84	0.89	0.99	0.87	0.89	1.00	0.86	0.36	0.66	0.35	0.39	0.72	0.35	0.46	0.92	0.36
8	0.87	0.97	0.84	0.88	0.98	0.85	0.88	0.99	0.85	0.40	0.66	0.39	0.43	0.72	0.40	0.50	0.92	0.40
9	0.86	0.96	0.84	0.88	0.97	0.85	0.88	0.99	0.85	0.44	0.66	0.42	0.47	0.72	0.43	0.52	0.92	0.43
10	0.85	0.94	0.84	0.87	0.96	0.84	0.89	0.99	0.85	0.47	0.66	0.45	0.49	0.72	0.46	0.55	0.92	0.46
11	0.84	0.92	0.83	0.86	0.95	0.84	0.86	0.97	0.84	0.49	0.66	0.48	0.52	0.72	0.48	0.57	0.92	0.49
12	0.83	0.91	0.84	0.87	0.94	0.86	0.85	0.98	0.84	0.53	0.66	0.50	0.55	0.72	0.50	0.61	0.92	0.51
13	0.83	0.90	0.82	0.85	0.93	0.85	0.84	0.97	0.84	0.55	0.66	0.50	0.58	0.72	0.51	0.63	0.92	0.51
14	0.82	0.91	0.83	0.86	0.94	0.85	0.84	0.98	0.82	0.56	0.66	0.50	0.58	0.72	0.51	0.62	0.92	0.52
15	0.81	0.94	0.82	0.85	0.96	0.85	0.84	0.97	0.82	0.53	0.66	0.51	0.56	0.72	0.52	0.60	0.92	0.52
16	0.81	0.97	0.82	0.85	0.97	0.84	0.83	0.99	0.83	0.49	0.66	0.51	0.52	0.72	0.52	0.57	0.92	0.52
17	0.90	0.99	0.83	0.90	0.99	0.86	0.84	0.99	0.81	0.46	0.66	0.51	0.49	0.72	0.52	0.54	0.92	0.52
18	0.96	1.00	0.84	0.91	0.99	0.86	0.85	1.00	0.84	0.44	0.66	0.52	0.46	0.72	0.52	0.53	0.92	0.52
19	0.98	1.00	0.84	0.94	1.00	0.86	0.88	0.99	0.84	0.43	0.66	0.52	0.46	0.72	0.52	0.51	0.92	0.53
20	0.99	1.00	0.85	0.95	1.00	0.85	0.86	1.00	0.85	0.43	0.66	0.52	0.46	0.72	0.53	0.51	0.92	0.53

Table B.6: Coverage and Length for 90 Confidence Interval with n=500 in Experiment 2

	GDP																	
	Coverage									Length								
	High			Medium			Low			High			Medium			Low		
h	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**	**FLP**	**FLP**$_j$	**LP**
1	0.92	1.00	0.87	0.91	1.00	0.87	0.94	1.00	0.89	0.23	0.54	0.18	0.24	0.65	0.18	0.29	0.89	0.19
2	0.92	1.00	0.86	0.89	1.00	0.88	0.89	1.00	0.87	0.22	0.54	0.22	0.24	0.65	0.22	0.32	0.89	0.22
3	0.89	0.99	0.87	0.89	1.00	0.85	0.90	1.00	0.86	0.26	0.54	0.26	0.29	0.65	0.26	0.36	0.89	0.26
4	0.89	0.99	0.85	0.89	0.99	0.85	0.89	1.00	0.86	0.29	0.54	0.28	0.32	0.65	0.29	0.39	0.89	0.29
5	0.89	0.98	0.85	0.85	0.99	0.85	0.89	1.00	0.86	0.31	0.54	0.30	0.33	0.65	0.31	0.41	0.89	0.31
6	0.89	0.99	0.86	0.87	0.99	0.86	0.90	1.00	0.86	0.31	0.54	0.31	0.34	0.65	0.31	0.41	0.89	0.32
7	0.88	0.99	0.87	0.86	0.99	0.83	0.86	0.99	0.85	0.32	0.54	0.32	0.35	0.65	0.32	0.42	0.89	0.32
8	0.86	0.99	0.85	0.85	0.99	0.85	0.86	1.00	0.85	0.33	0.54	0.31	0.36	0.65	0.31	0.43	0.89	0.32
9	0.86	0.99	0.86	0.84	0.99	0.83	0.85	1.00	0.85	0.32	0.54	0.31	0.35	0.65	0.31	0.42	0.89	0.32
10	0.88	0.99	0.86	0.86	1.00	0.85	0.84	1.00	0.84	0.31	0.54	0.31	0.34	0.65	0.32	0.41	0.89	0.32
11	0.92	0.99	0.87	0.88	1.00	0.85	0.85	1.00	0.86	0.29	0.54	0.32	0.33	0.65	0.32	0.39	0.89	0.32
12	0.95	1.00	0.84	0.88	1.00	0.85	0.86	1.00	0.85	0.29	0.54	0.32	0.32	0.65	0.32	0.41	0.89	0.32
13	0.96	1.00	0.87	0.91	1.00	0.87	0.89	1.00	0.86	0.28	0.54	0.32	0.31	0.65	0.32	0.39	0.89	0.32
14	0.98	1.00	0.86	0.89	1.00	0.88	0.88	1.00	0.87	0.28	0.54	0.32	0.30	0.65	0.32	0.40	0.89	0.32
15	0.98	1.00	0.88	0.90	1.00	0.87	0.83	1.00	0.88	0.27	0.54	0.32	0.30	0.65	0.32	0.38	0.89	0.33
16	0.99	1.00	0.87	0.90	1.00	0.88	0.84	1.00	0.86	0.27	0.54	0.32	0.30	0.65	0.32	0.38	0.89	0.33
17	0.99	1.00	0.88	0.89	1.00	0.86	0.81	1.00	0.85	0.27	0.54	0.32	0.30	0.65	0.32	0.37	0.89	0.33
18	1.00	1.00	0.88	0.86	1.00	0.88	0.82	1.00	0.85	0.27	0.54	0.32	0.30	0.65	0.32	0.39	0.89	0.33
19	1.00	1.00	0.87	0.87	1.00	0.87	0.81	1.00	0.87	0.27	0.54	0.32	0.31	0.65	0.32	0.39	0.89	0.33
20	1.00	1.00	0.89	0.87	1.00	0.87	0.81	1.00	0.88	0.27	0.54	0.32	0.30	0.65	0.32	0.39	0.89	0.33

Figure B.1: IRF to a Monetary Policy shock with High Smoothness and T=500 in Experiment 1

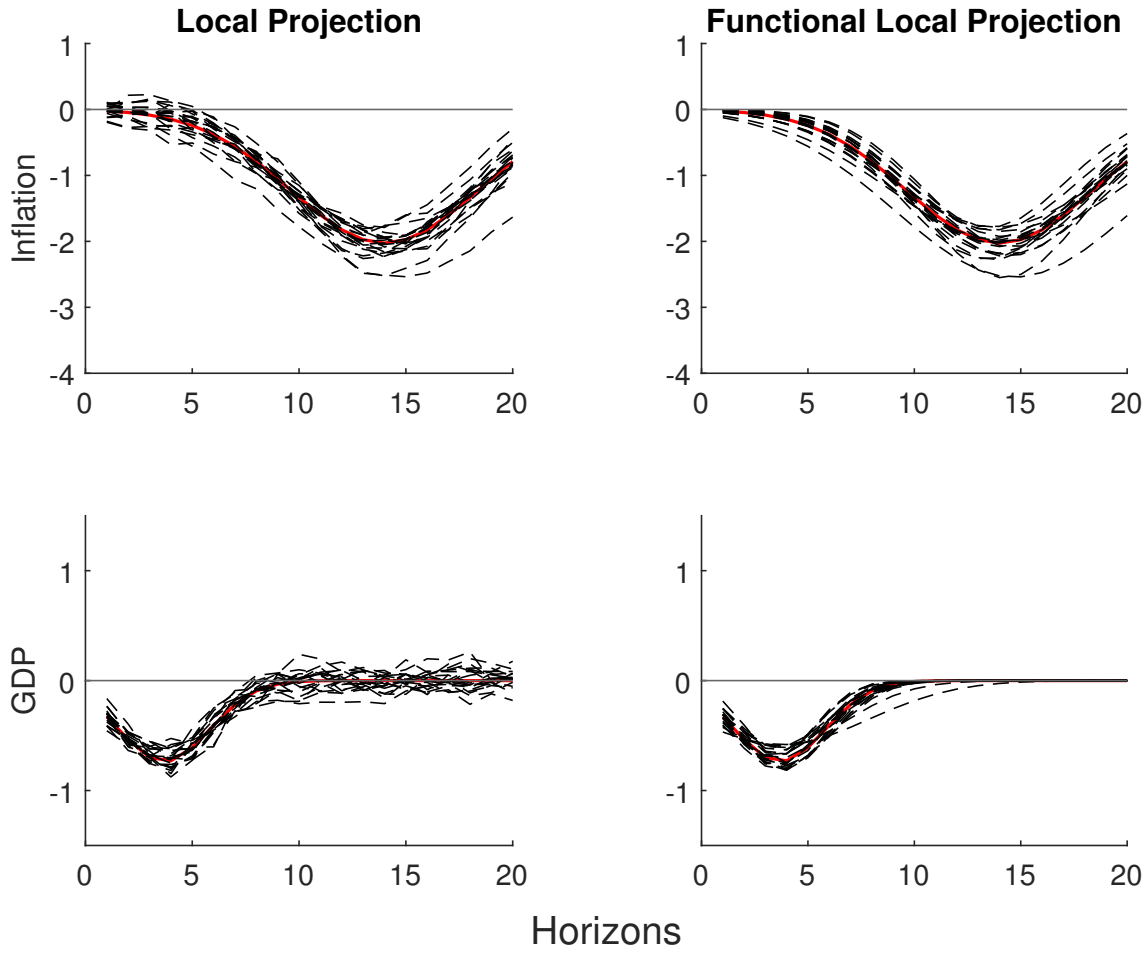


Figure B.2: IRF to a Monetary Policy shock with Medium Smoothness and n=500 in Experiment 1

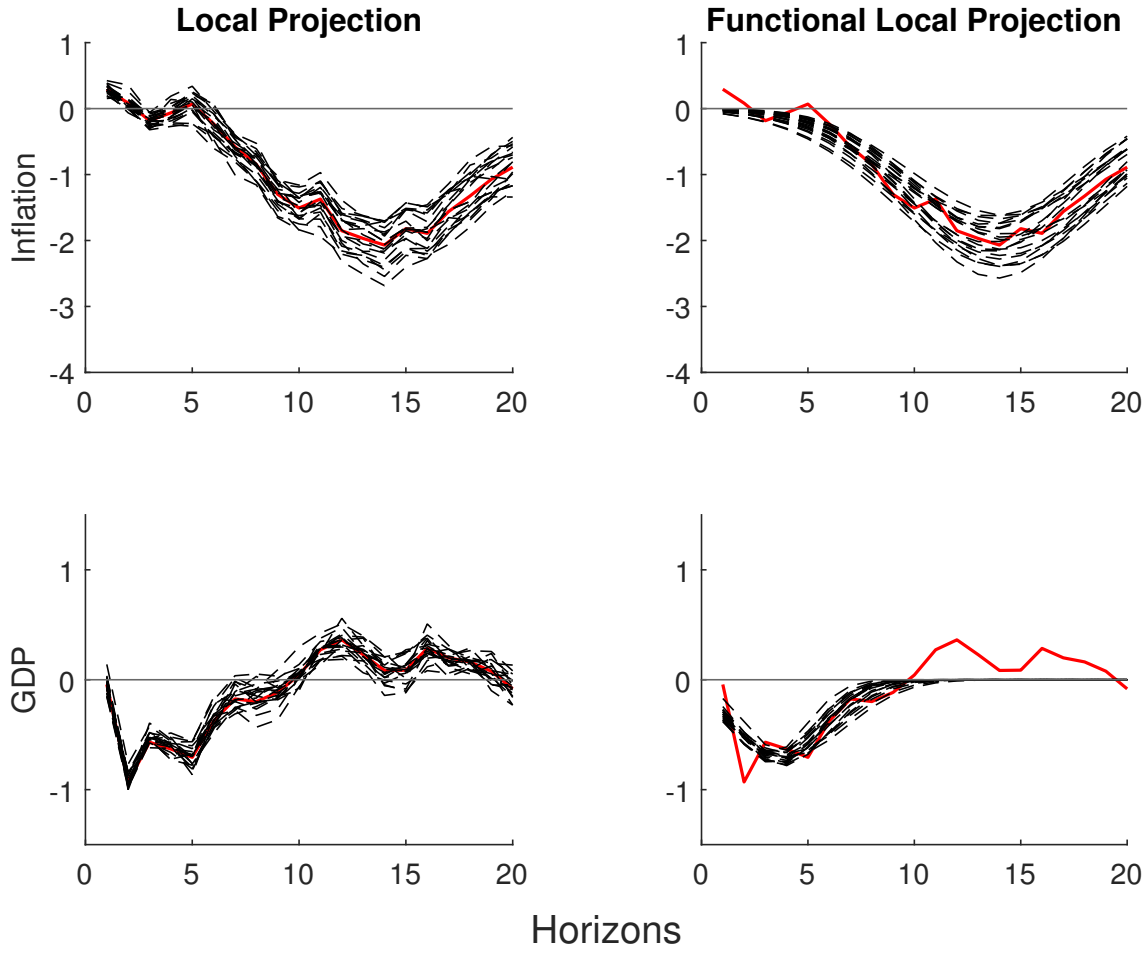


Figure B.3: IRF to a Monetary Policy shock with Low Smoothness and n=500 in Experiment 1

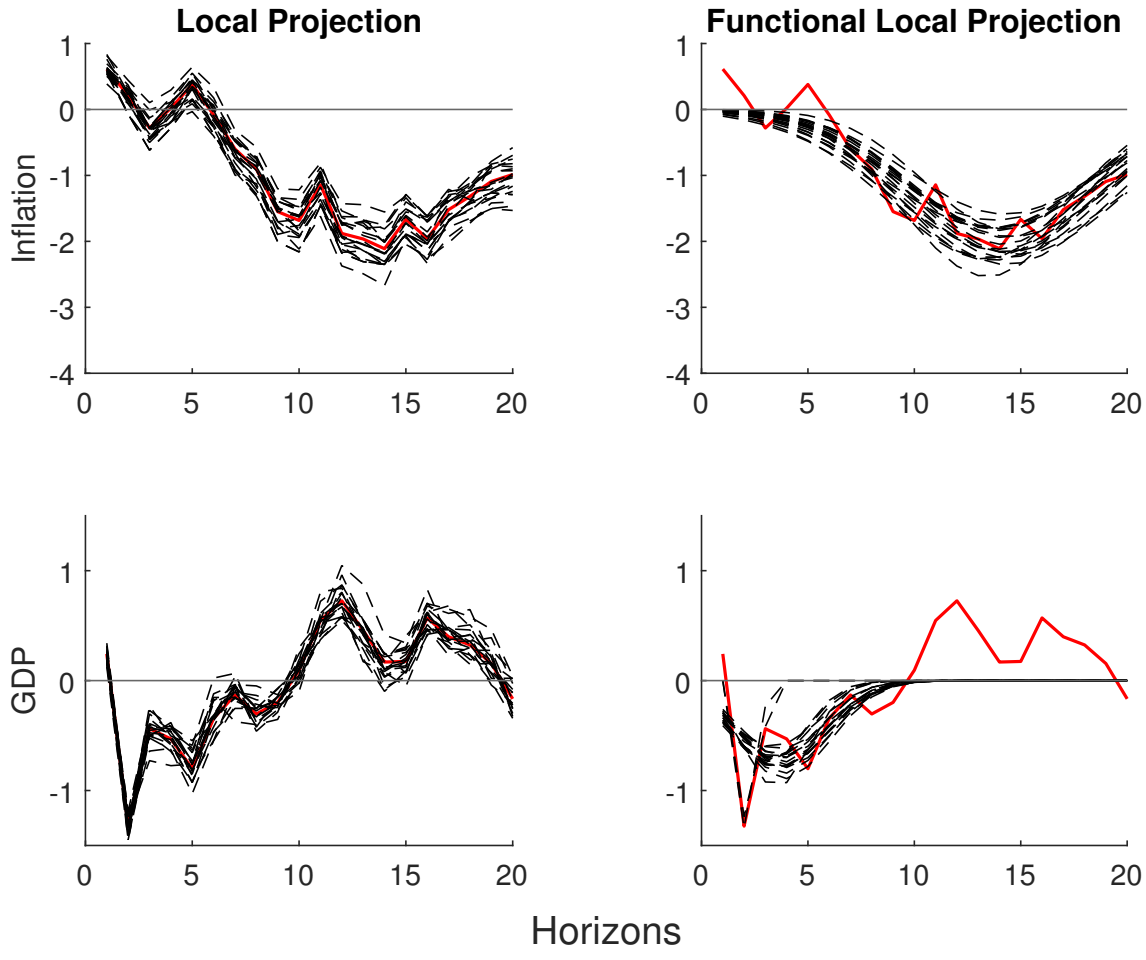


Figure B.4: Confidence Interval Simulations with High Smoothness and n=500 in Experiment 1

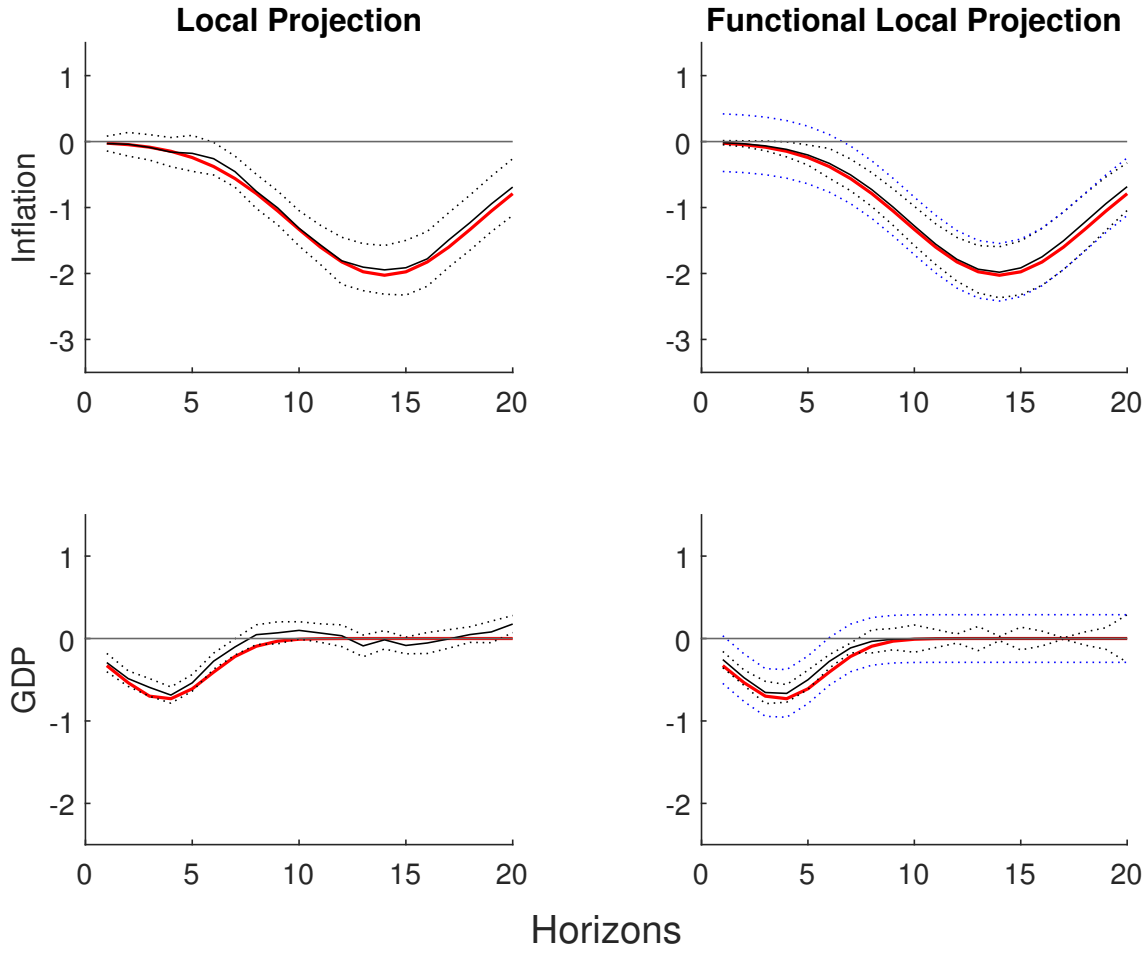


Figure B.5: Confidence Interval Simulations with Medium Smoothness and n=500 in Experiment 1

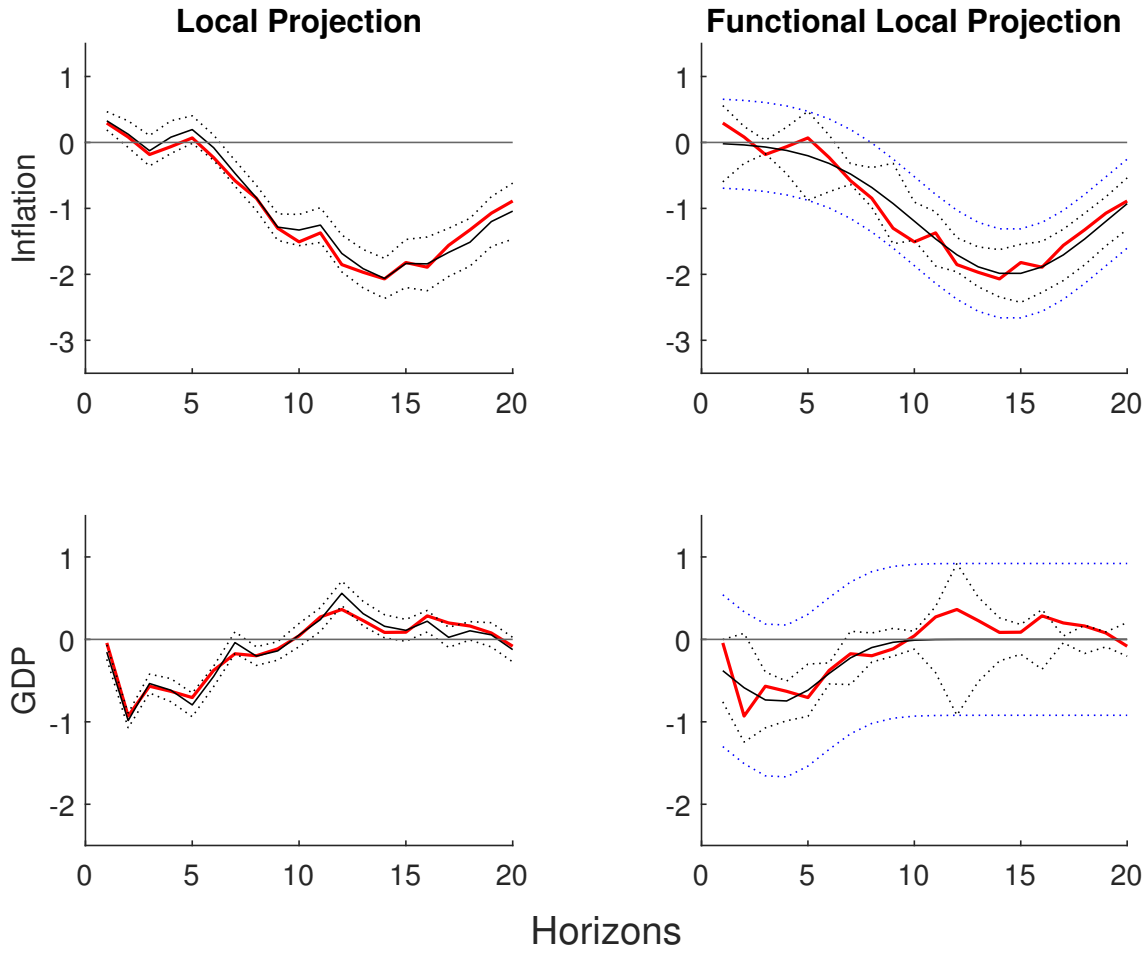


Figure B.6: Confidence Interval Simulations with Low Smoothness and n=500 in Experiment 1

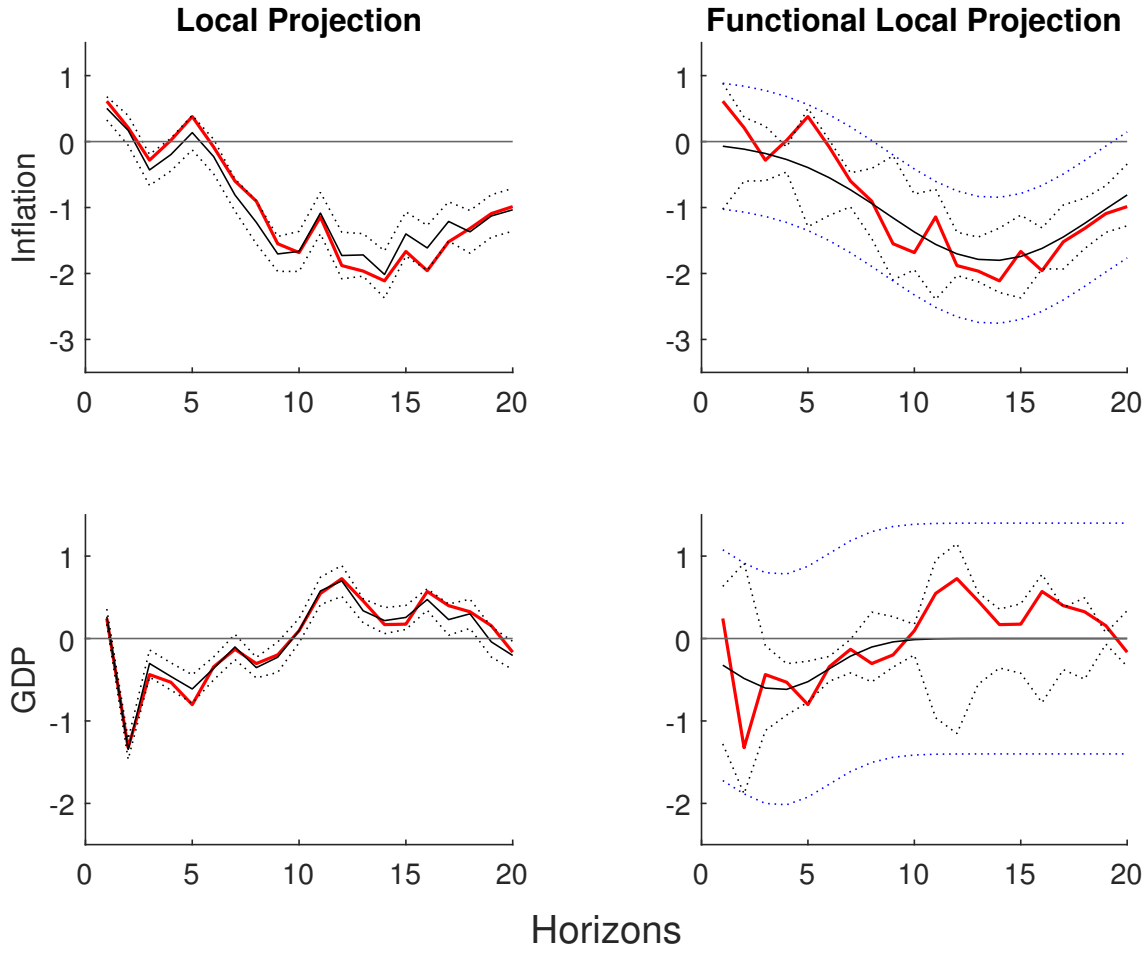


Figure B.7: Histogram of a,b,c with the High Smoothness and n=500 in Experiment 1

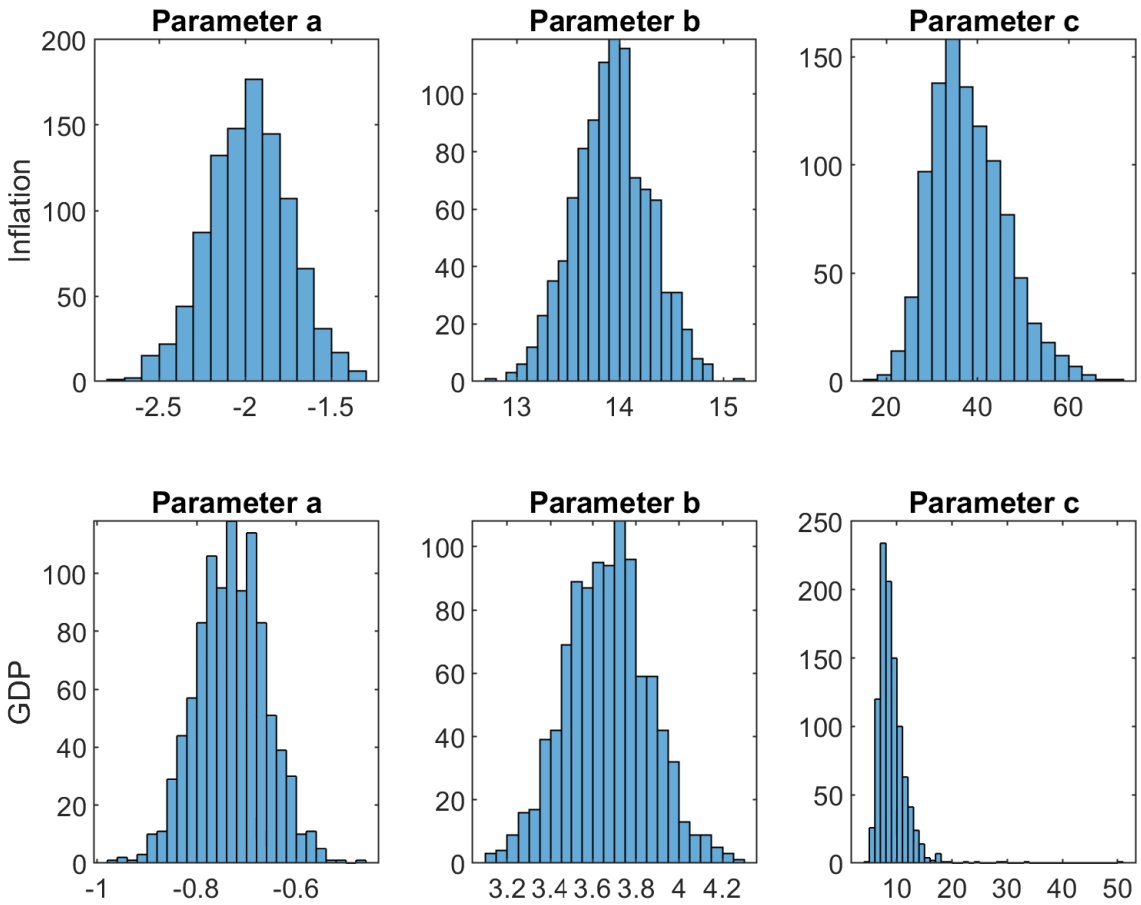


Figure B.8: Histogram of a,b,c with the Medium Smoothness and n=500 in Experiment 1

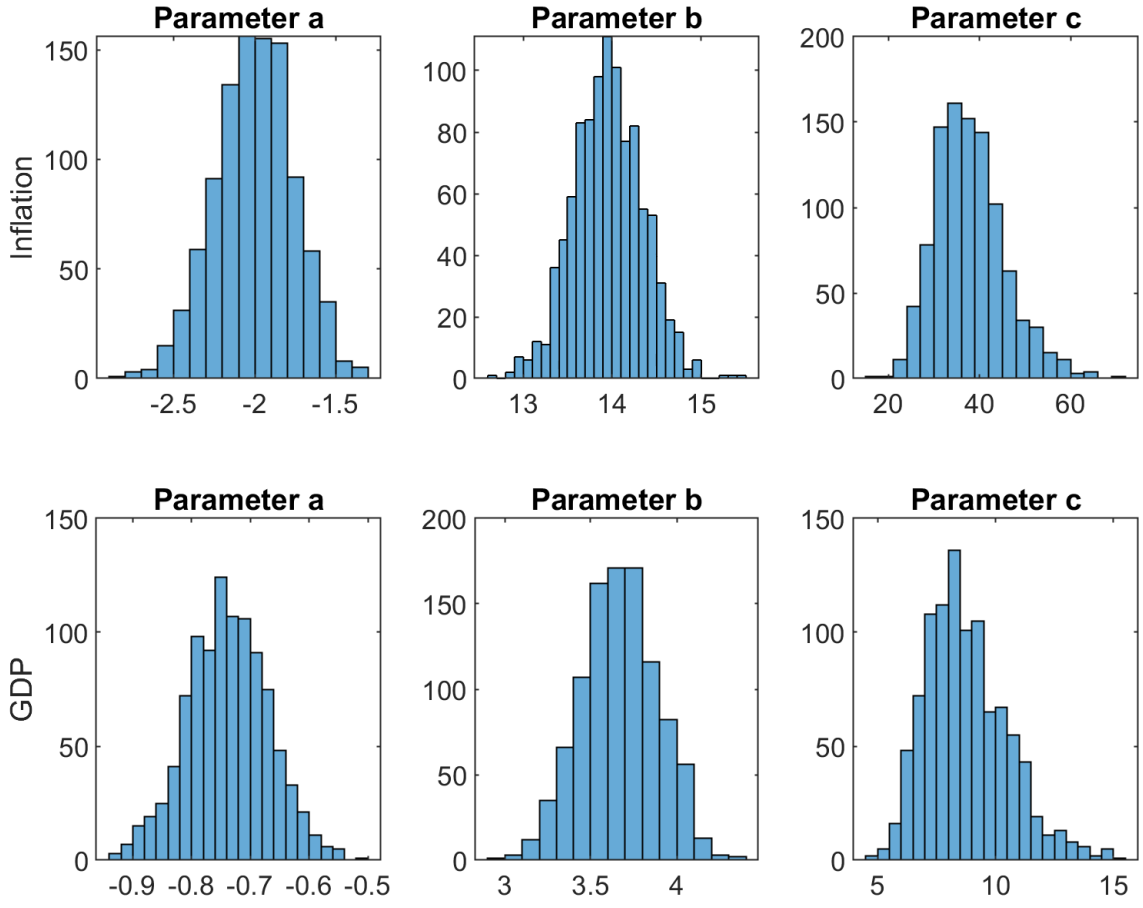


Figure B.9: Histogram of a,b,c with the Low Smoothness and n=500 in Experiment 1

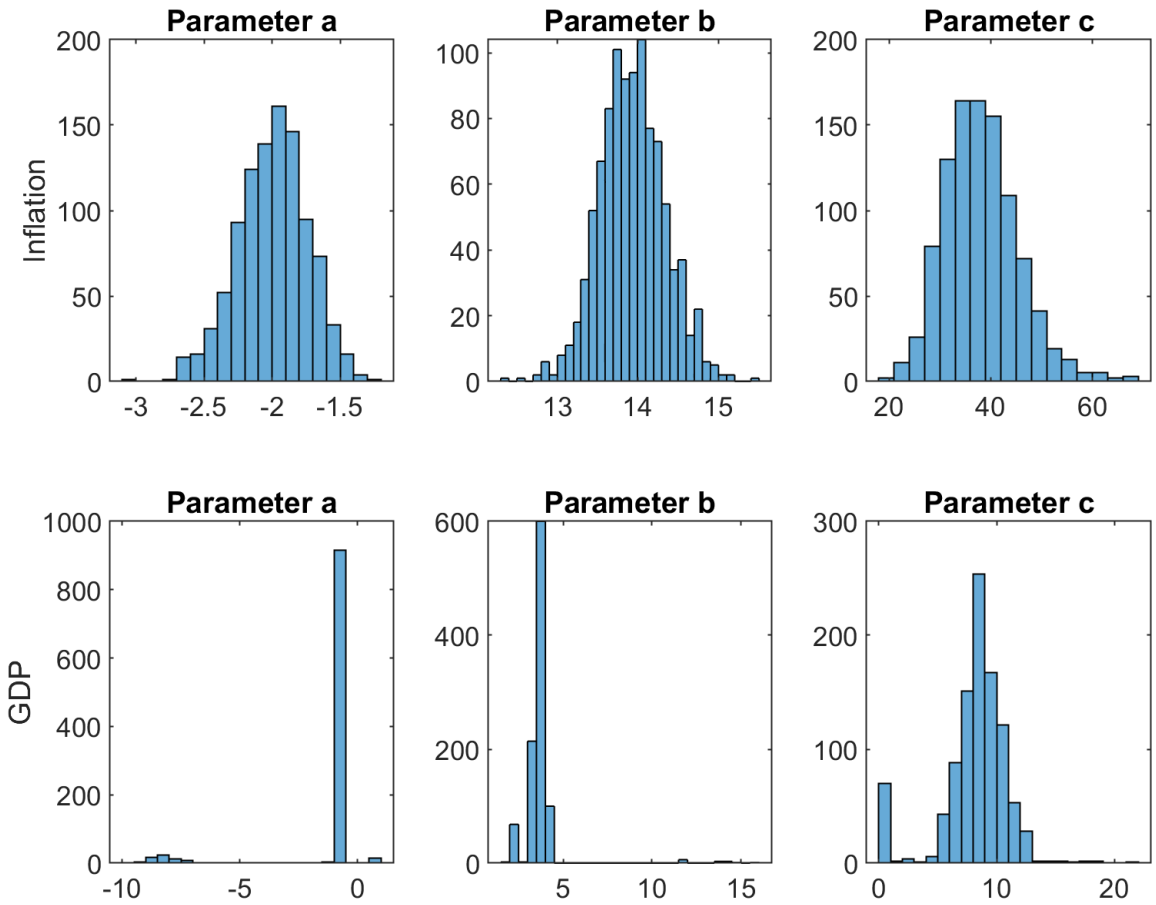


Figure B.10: Cumulative IRFs to a TFP shock - Observed Shock Specification - 2

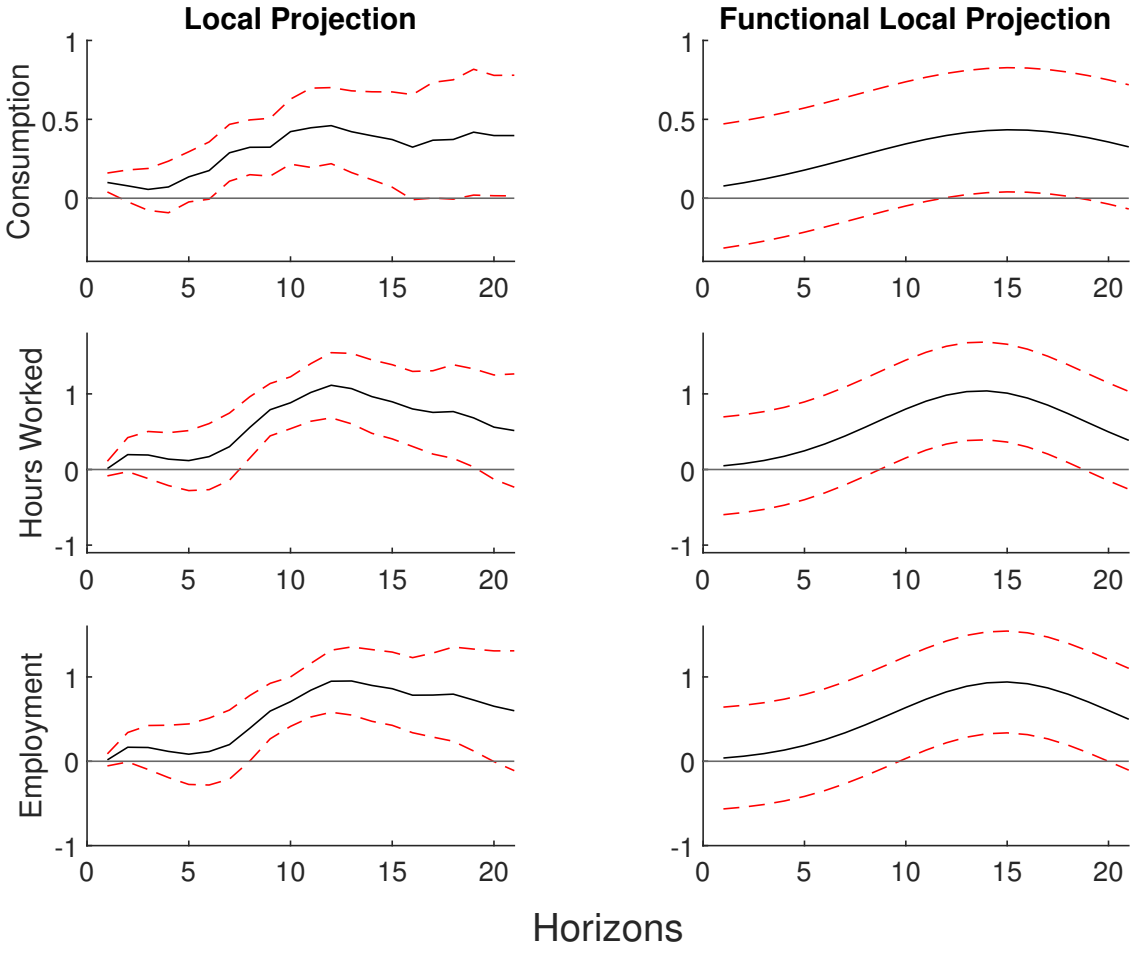


Figure B.11: Cumulative IRFs to a TFP shock - IV Shock Specification - 2

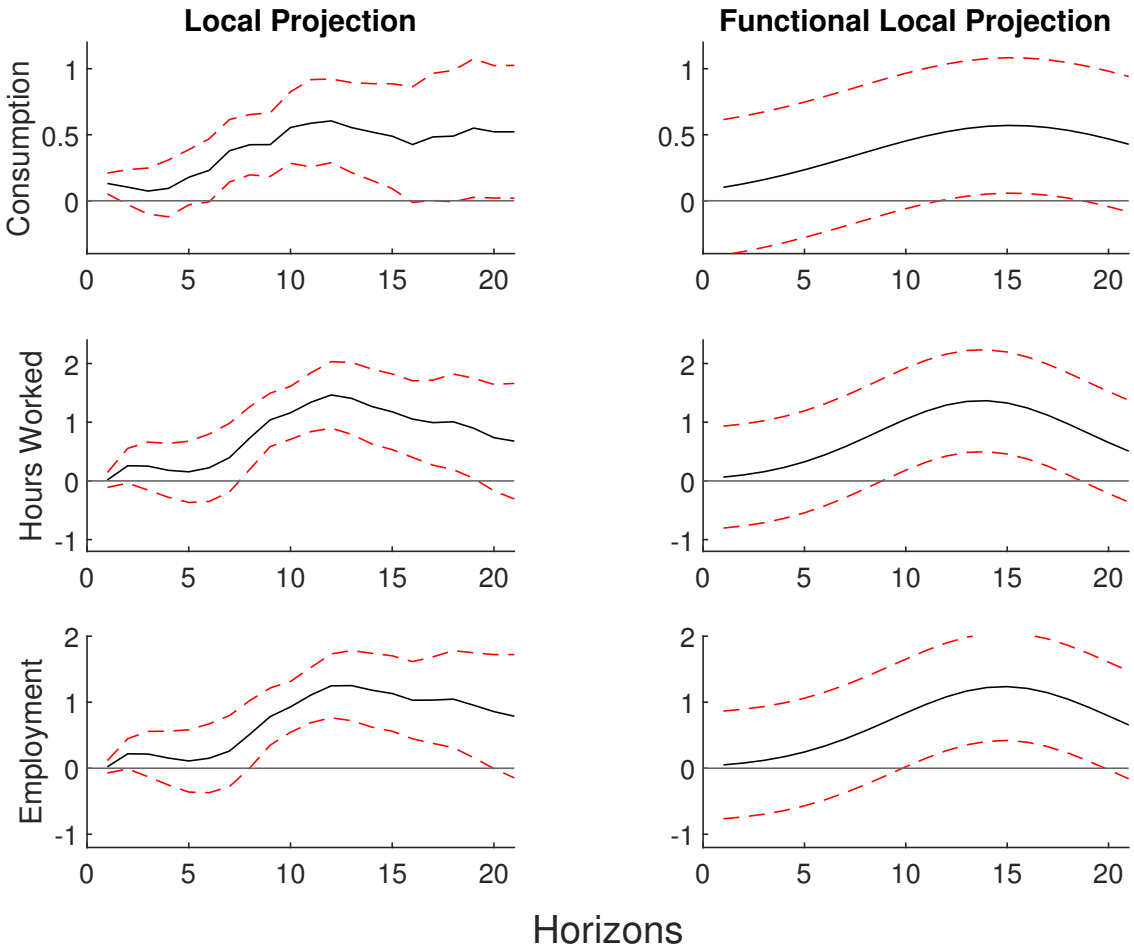


Figure B.12: Cumulative IRFs to a TFP shock - IV Shock +TV Specification - 2

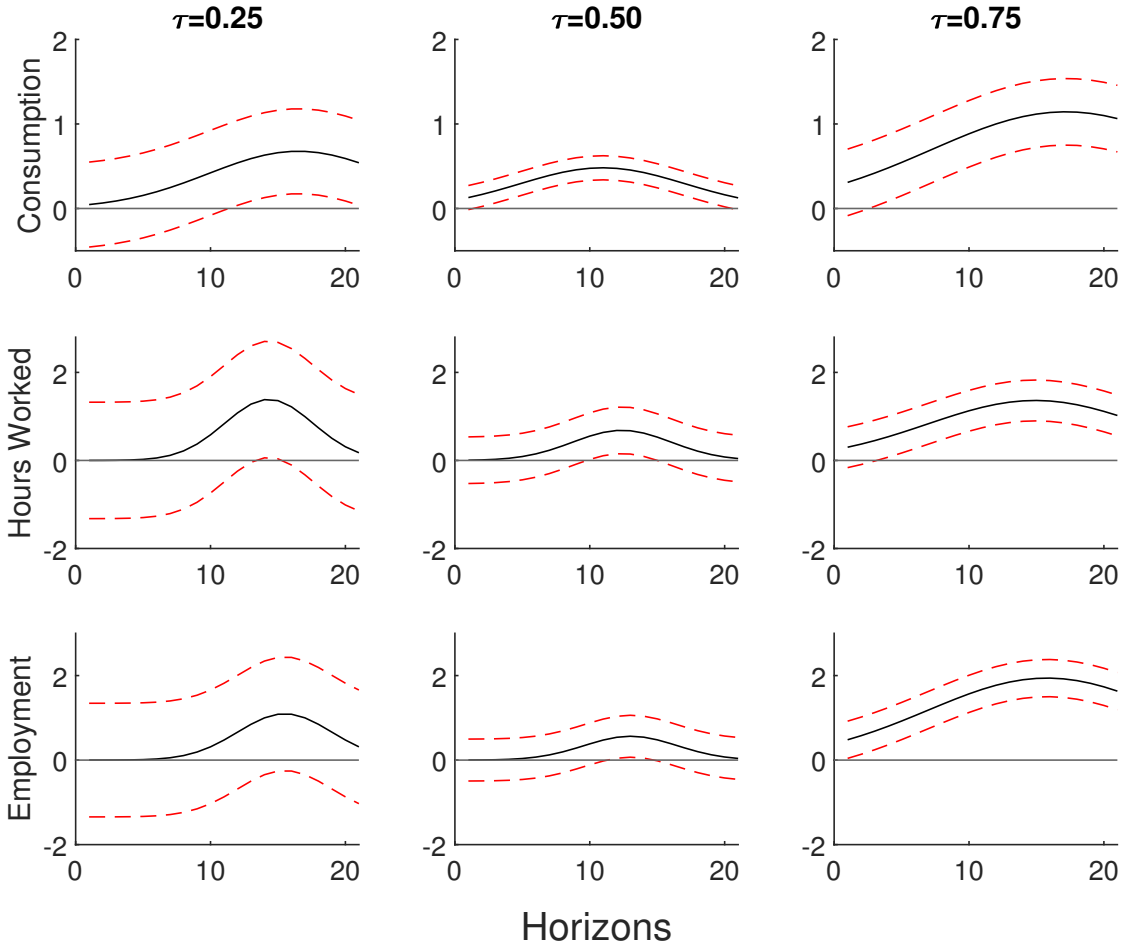


Figure B.13: Non-Cumulative IRFs to a TFP shock - Observed Shock Specification - 1

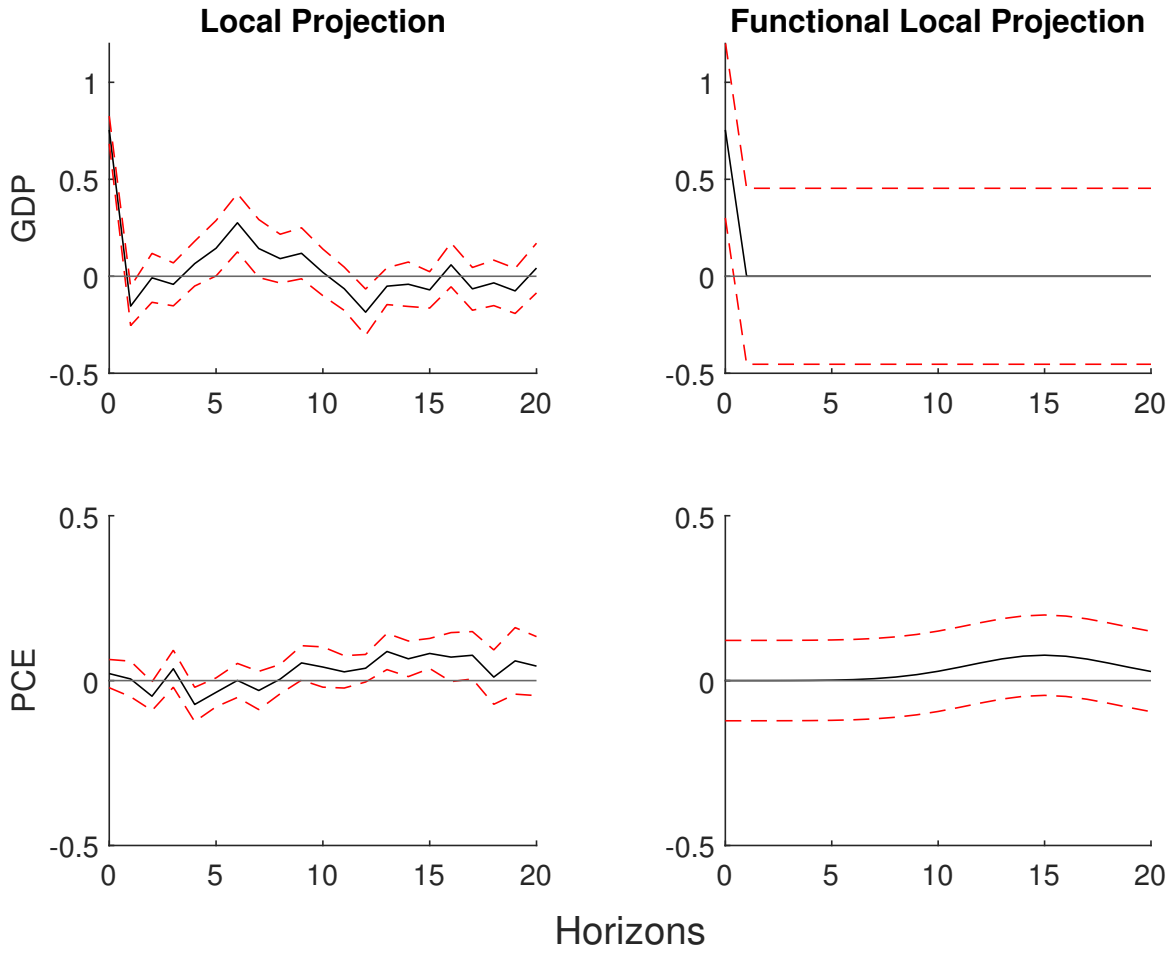


Figure B.14: Non-Cumulative IRFs to a TFP shock - Observed Shock Specification - 2

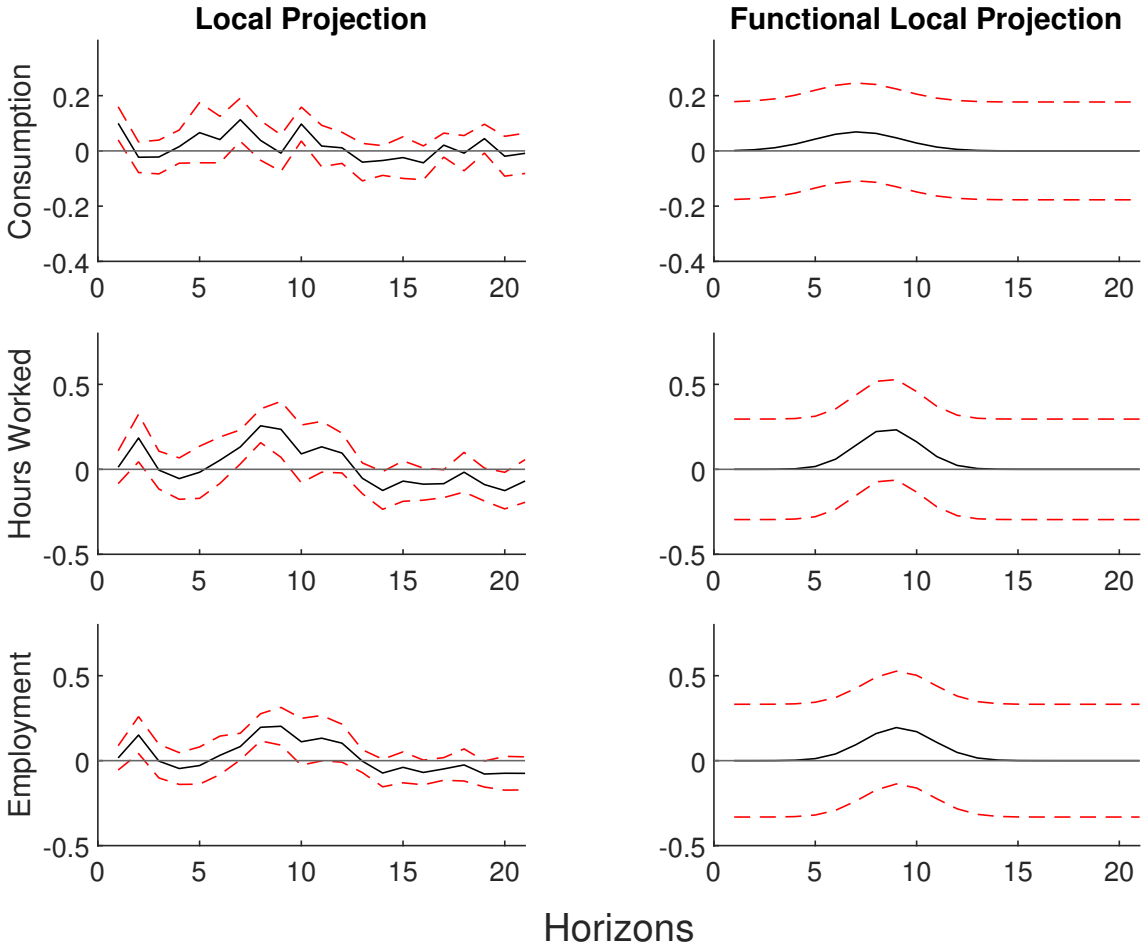


Figure B.15: Non-Cumulative IRFs to a TFP shock - IV Shock Specification - 1

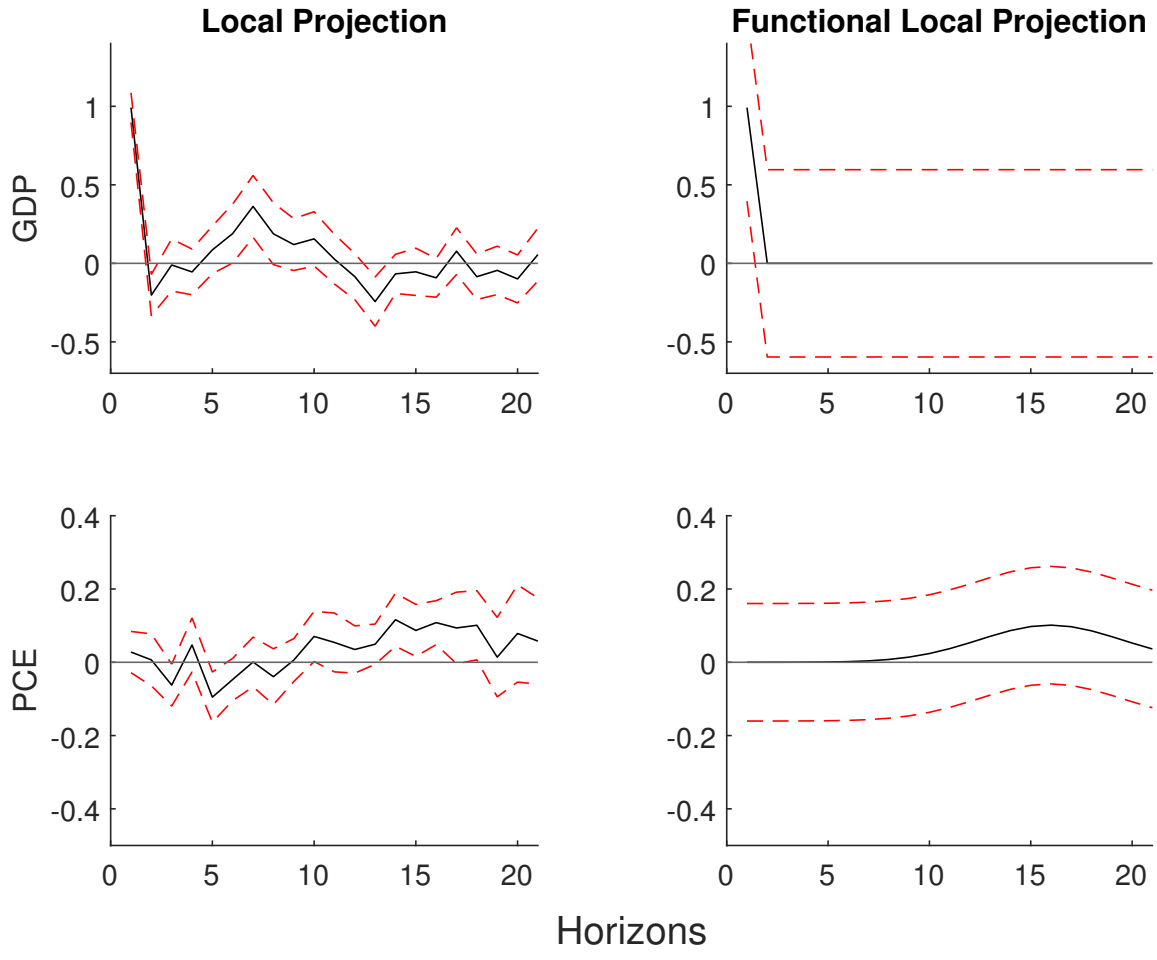


Figure B.16: Non-Cumulative IRFs to a TFP shock - IV Shock Specification - 2

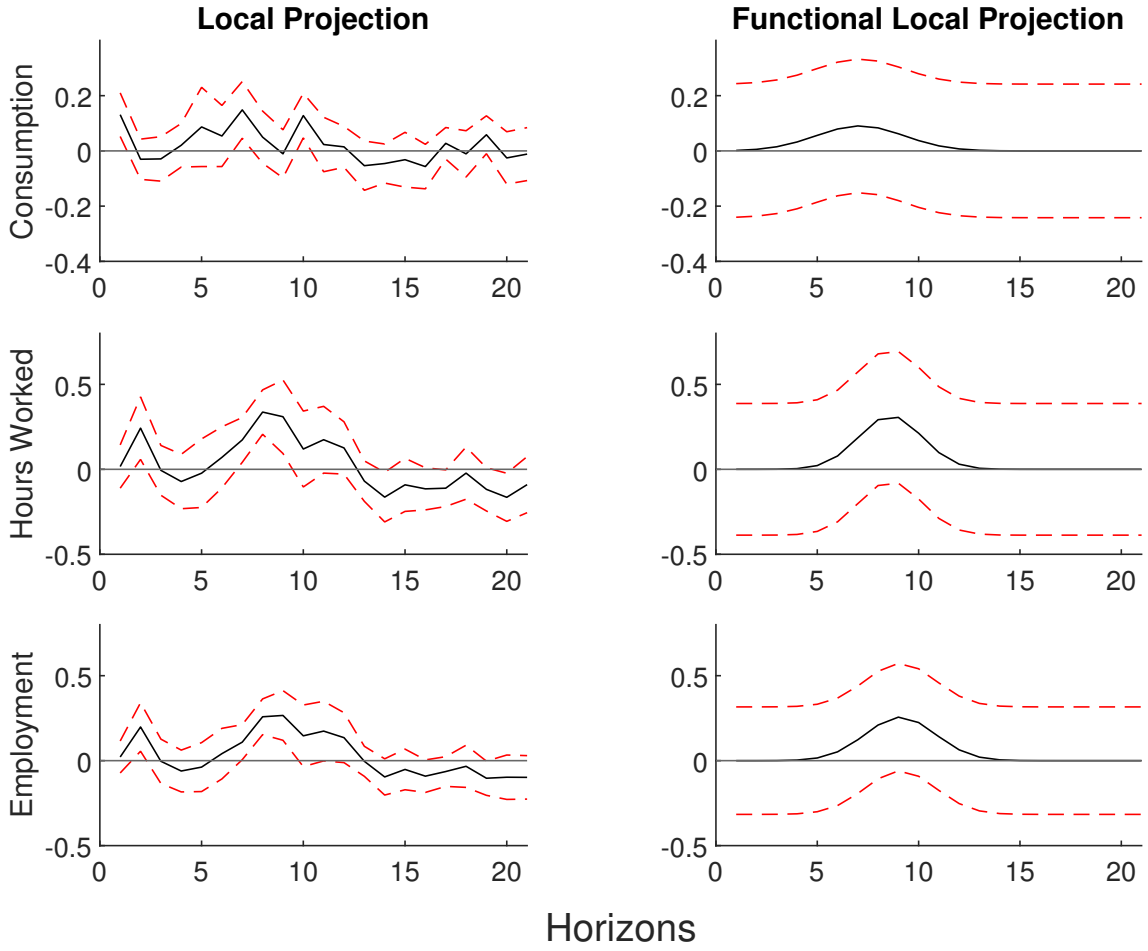


Figure B.17: Non-Cumulative IRFs to a TFP shock - IV Shock +TV Specification - 1

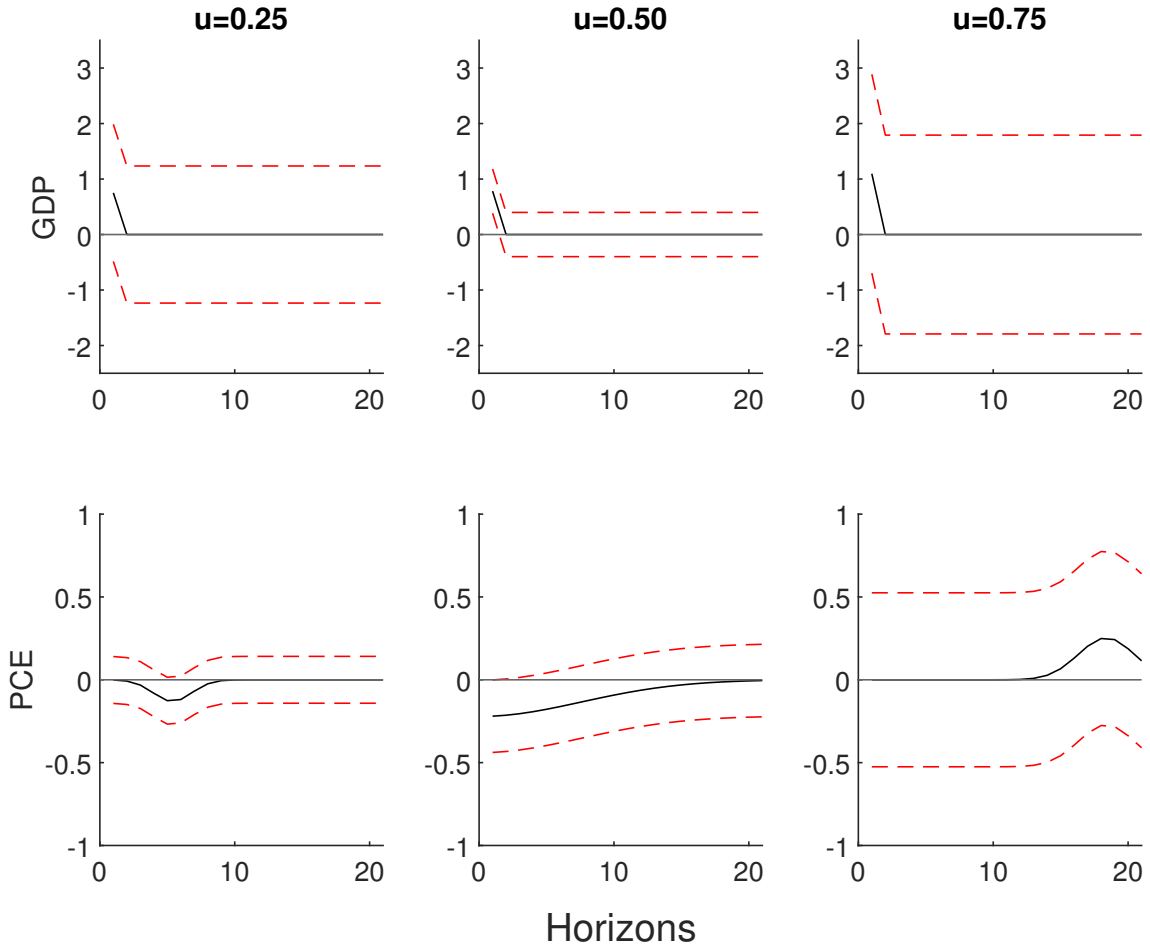
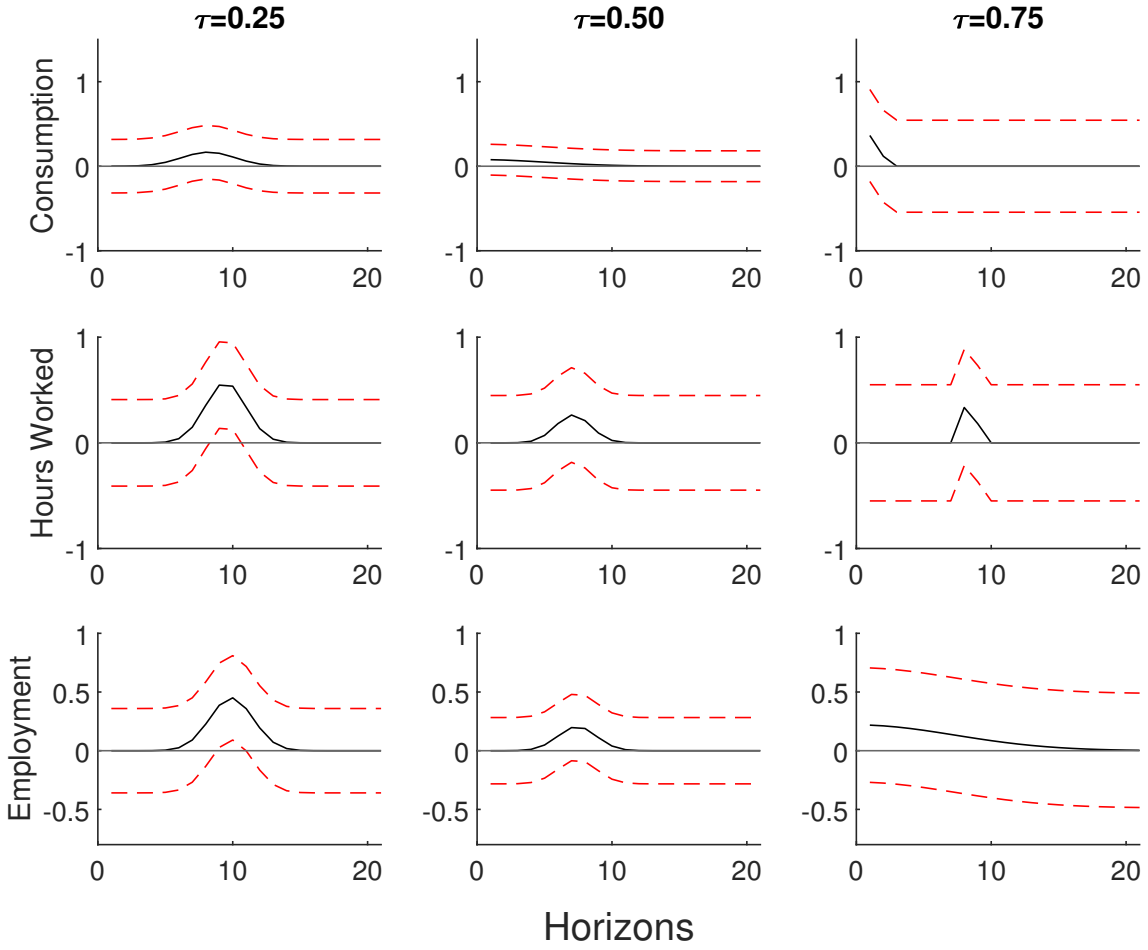


Figure B.18: Non-Cumulative IRFs to a TFP shock - IV Shock +TV Specification - 2



B.5 Tables

Table B.7: Estimated Parameters of the Gaussian basis for cumulative IRFs

		GDP	Inflation	Consum	HoursW	Emp
Direct Shock	a	1.10	-1.05	0.33	0.53	0.46
	b	14.20	15.39	15.62	14.32	15.20
	c	138.88	164.84	124.35	18.12	20.24
IV	a	1.45	-1.39	0.43	0.69	0.60
	b	14.20	15.39	15.62	14.32	15.20
	c	138.88	164.84	124.35	18.12	20.24
IV,$\tau = 0.25$	a	1.66	-0.43	0.68	1.39	1.10
	b	13.85	9.55	16.51	14.32	15.50
	c	86.71	31.21	89.70	21.51	24.16
IV,$\tau = 0.50$	a	1.41	-2.02	0.48	0.68	0.56
	b	11.47	17.35	10.95	12.39	13.03
	c	125.22	200.00	75.21	26.70	23.79
IV,$\tau = 0.75$	a	1.81	-1.30	1.14	1.36	1.94
	b	14.58	11.11	17.18	14.92	15.79
	c	200.00	54.23	200.00	128.26	156.87

Note: This table displays the point estimates of the parameters a,b,c which describe the Gaussian basis function under the FLP framework. Each column represents the response variables to a TFP shocks. The rows exhibits each specification considered in the empirical study. The last three rows correspond to the IV case with time-varying parameters at the first, second and third quantiles of the sample.

Table B.8: Estimated Parameters of the Gaussian basis for non cumulative IRFs

Specification		GDP	Inflation	Consum.	Hours.W	Emp.
Direct Shock	a	1.10	-0.14	0.05	0.19	0.17
	b	0.81	4.15	8.00	9.45	9.91
	c	0.06	20.50	9.82	6.48	6.60
IV	a	1.29	-0.18	0.07	0.25	0.22
	b	0.84	4.15	8.00	9.45	9.91
	c	0.06	20.50	9.82	6.48	6.60
IV, $\tau = 0.25$	a	1.07	-0.13	0.17	0.57	0.45
	b	0.85	5.39	8.19	9.46	9.90
	c	0.06	4.03	7.84	4.46	5.24
IV, $\tau = 0.50$	a	1.47	-0.22	0.08	0.26	0.21
	b	0.82	0.00	0.00	7.12	7.42
	c	0.05	114.25	51.05	3.38	4.05
IV, $\tau = 075$	a	1.28	0.25	1.35	9.97	0.22
	b	0.91	18.39	1.42	8.48	0.00
	c	0.05	8.61	0.14	0.07	106.01

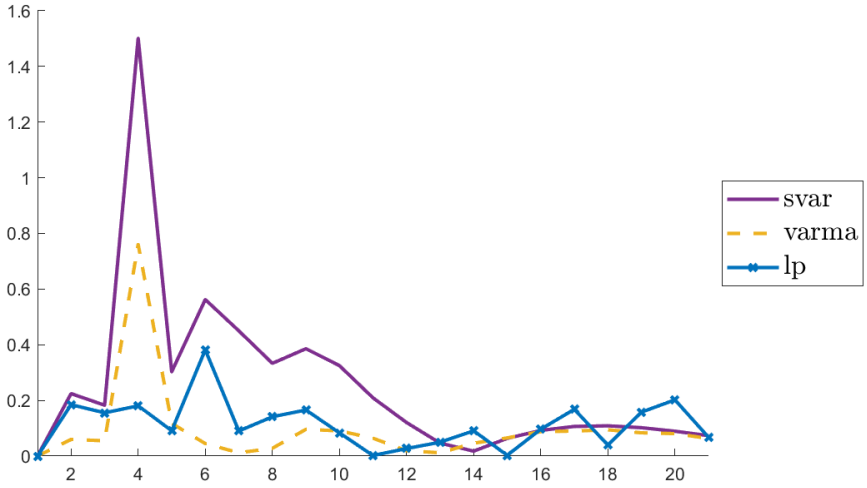
Note: This table displays the point estimates of the parameters a,b,c which describe the Gaussian basis function under the FLP framework. Each column represents the response variables to a TFP shocks. The rows exhibits each specification considered in the empirical study. The last three rows correspond to the IV case with time-varying parameters at the first, second and third quantiles of the sample.

Appendix C

CHAPTER 3

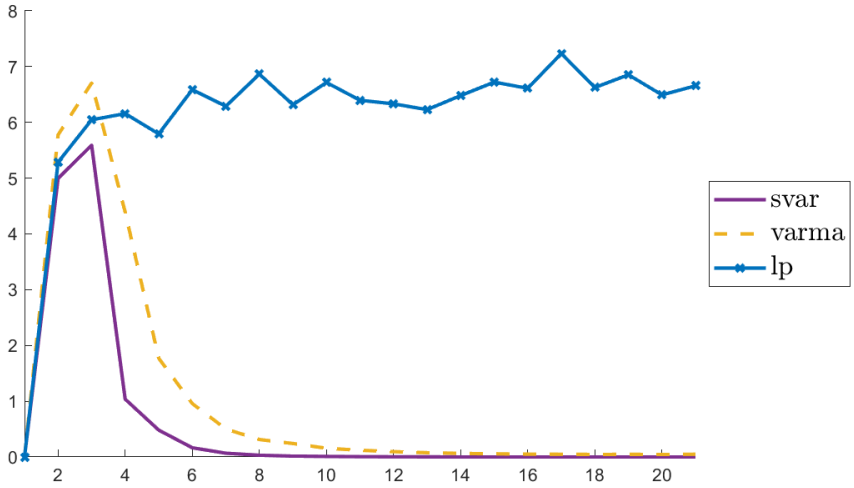
C.1 Plots

Figure C.1: Recursive MP: Mean Bias of Estimators



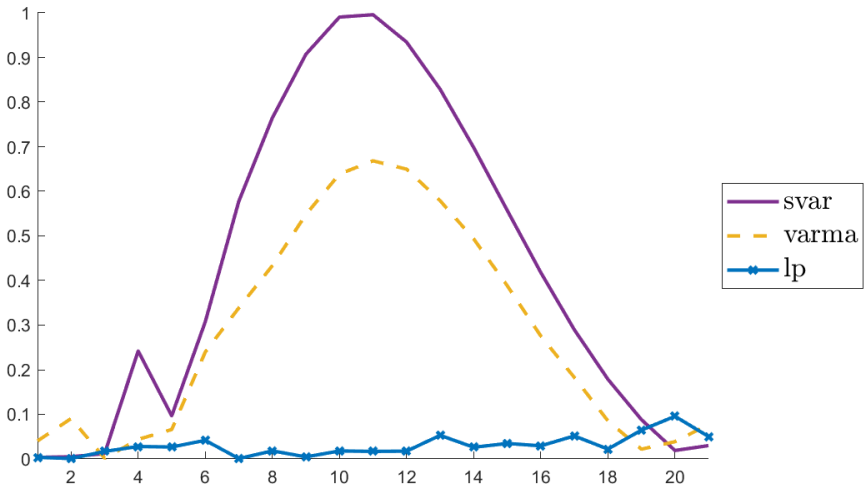
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.2: Recursive MP: Standard Deviation of Estimators



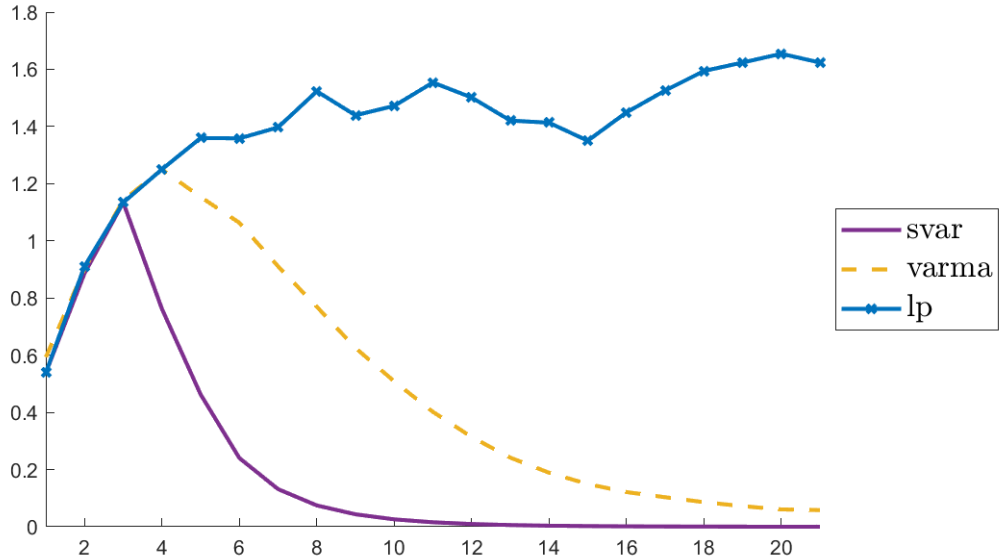
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.3: Observed Shock G: Mean Bias of Estimators



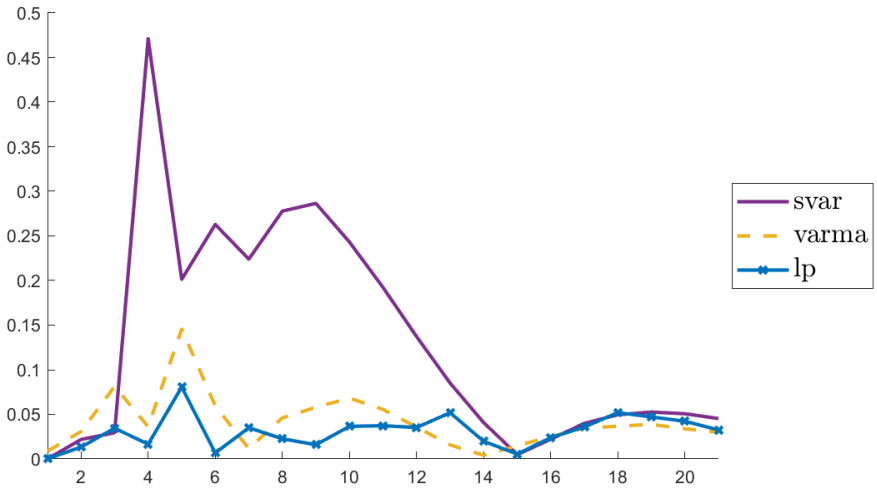
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.4: Observed Shock G: Standard Deviation of Estimators



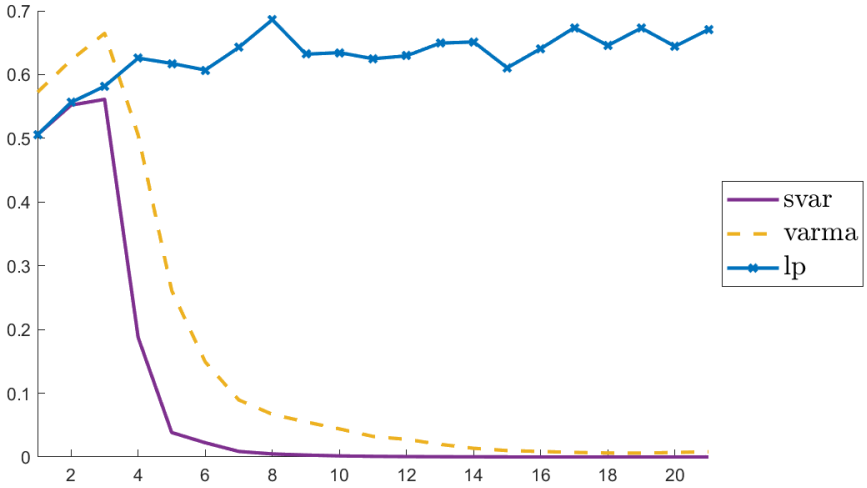
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.5: Observed Shock MP: Mean Bias of Estimators



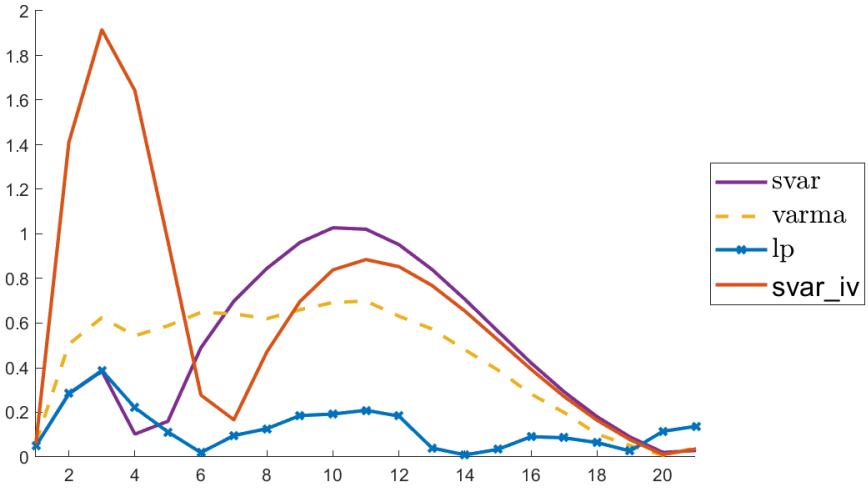
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.6: Observed Shock MP: Standard Deviation of Estimators



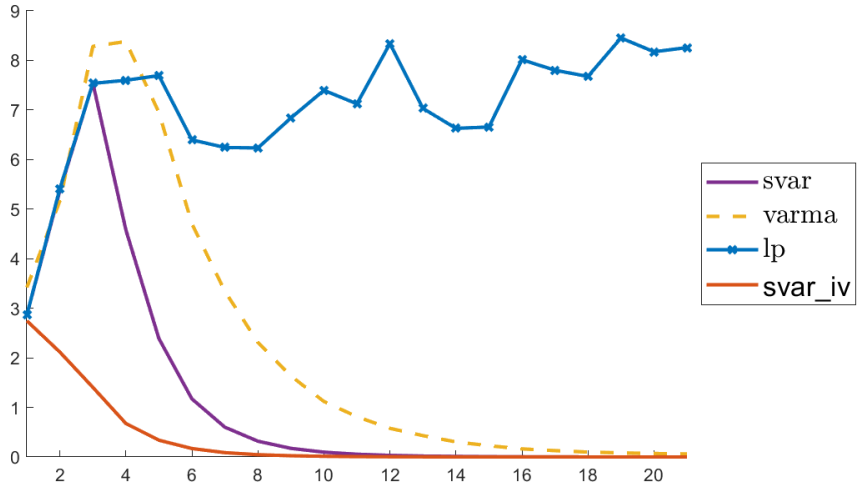
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.7: IV G: Mean Bias of Estimators



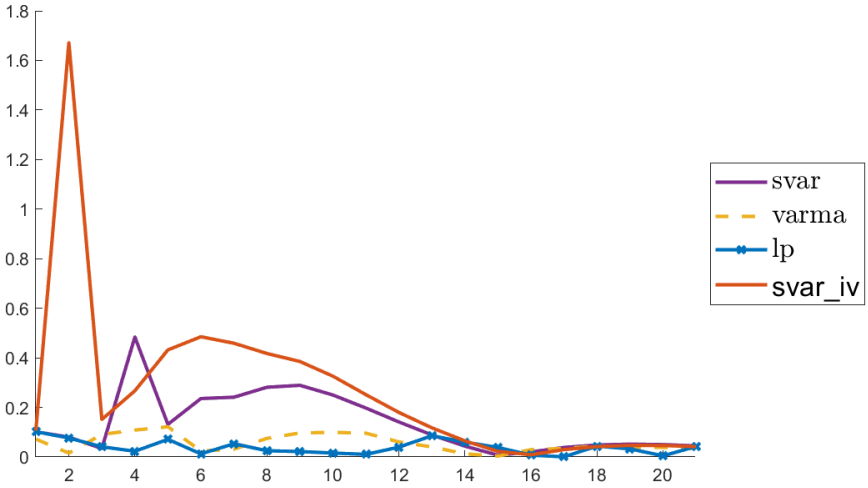
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.8: IV Shock G: Standard Deviation of Estimators



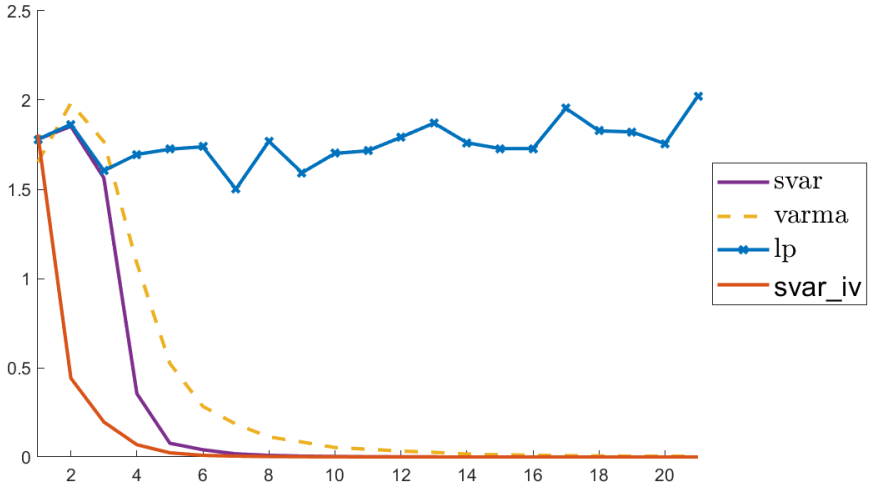
Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.9: IV Shock MP: Mean Bias of Estimators



Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Figure C.10: IV Shock MP: Standard Deviation of Estimators



Note: IRFs have been divided by the root mean squared value of the true IRFs out to horizon 20 in order to cancel out units of the response variables. For each of the 1000 Monte Carlo replications, one specification is considered.

Bibliography

Alexopoulos, M. (2011). Read all about it!! what happens following a technology shock? *American Economic Review*, 101(4):1144–79.

Anderson, J. E. and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American economic review*, 93(1):170–192.

Ang, A. and Chen, J. (2007). Capm over the long run: 1926–2001. *Journal of Empirical Finance*, 14(1):1–40.

Athanasopoulos, G., Poskitt, D. S., and Vahid, F. (2012). Two canonical varma forms: Scalar component models vis-à-vis the echelon form. *Econometric Reviews*, 31(1):60–83.

Athanasopoulos, G. and Vahid, F. (2008a). A complete varma modelling methodology based on scalar components. *Journal of Time Series Analysis*, 29(3):533–554.

Athanasopoulos, G. and Vahid, F. (2008b). Varma versus var for macroeconomic forecasting. *Journal of Business & Economic Statistics*, 26(2):237–252.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.

Barnichon, R. and Brownlees, C. (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, 101(3):522–530.

Barnichon, R. and Matthes, C. (2018). Functional Approximation of Impulse Responses. *Journal of Monetary Economics*.

Basu, S., Fernald, J. G., and Kimball, M. S. (2006). Are technology improvements contractionary? *American Economic Review*, 96(5):1418–1448.

Benati, L., Chan, J., Eisenstat, E., and Koop, G. (2020). Identifying noise shocks. *Journal of Economic Dynamics and Control*, 111:103780.

Blanchard, O. and Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics*, 117(4):1329–1368.

Boldea, O., Cornea-Madeira, A., and Hall, A. R. (2019). Bootstrapping structural change tests. *Journal of Econometrics*, 213(2):359–397.

Bruno, V. and Shin, H. S. (2015). Capital flows and the risk-taking channel of monetary policy. *Journal of monetary economics*, 71:119–132.

Cai, Z. (2002). A two–stage approach to additive time series models. *Statistica Neerlandica*, 56(4):415–433.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Cai, Z. and Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, pages 1321–1342.

Cai, Z. and Tiwari, R. C. (2000). Application of a local linear autoregressive model to bod time series. *Environmetrics: The official journal of the International Environmetrics Society*, 11(3):341–350.

Chan, J. C. and Eisenstat, E. (2017). Efficient estimation of bayesian varmas with time-varying coefficients. *Journal of Applied Econometrics*, 32(7):1277–1297.

Chan, J. C., Eisenstat, E., and Koop, G. (2016). Large bayesian varmas. *Journal of Econometrics*, 192(2):374–390.

Chan, J. C., Eisenstat, E., and Koop, G. (2022). Choosing between identification schemes in noisy-news models. *Studies in Nonlinear Dynamics & Econometrics*, 26(1):99–136.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2008). Are structural vars with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics*, 55(8):1337–1352.

Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.

Chen, B., Choi, J., and Escanciano, J. C. (2017). Testing for fundamental vector moving average representations. *Quantitative Economics*, 8(1):149–180.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.

Chen, B. and Hong, Y. (2016). Detecting for smooth structural changes in garch models. *Econometric Theory*, 32(3):740.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, 1:65–148.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics*, 8(2):262 – 302. Monetary Policy and Learning.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, 4:1–32.

Cooley, T. F. and Dwyer, M. (1998). Business cycle analysis without much theory a look at structural vars. *Journal of econometrics*, 83(1-2):57–88.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier.

Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R., Rao, S. S., et al. (2006). Statistical inference for time-varying arch processes. *Annals of Statistics*, 34(3):1075–1114.

Dahlhaus, R., Richter, S., Wu, W. B., et al. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013–1044.

Das, S. (2003). Modelling money, price and output in india: a vector autoregressive and moving average (varma) approach. *Applied Economics*, 35(10):1219–1225.

Dias, G. F. and Kapetanios, G. (2018). Estimation and forecasting in vector autoregressive moving average models for rich datasets. *Journal of Econometrics*, 202(1):75–91.

Dufour, J.-M. and Pelletier, D. (2022). Practical methods for modeling weak varma processes: identification, estimation and specification with a macroeconomic application. *Journal of Business & Economic Statistics*, 40(3):1140–1152.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Fan, J., Zhang, W., et al. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 27(5):1491–1518.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco.

Fernald, J. G. (2012). A quarterly, utilization-adjusted series on total factor productivity. Working Paper Series 2012-19, Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., and Watson, M. W. (2007). Abcs (and ds) of understanding vars. *American economic review*, 97(3):1021–1026.

Francis, N., Owyang, M. T., Roush, J. E., and DiCecio, R. (2014). A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks. *Review of Economics and Statistics*, 96(4):638–647.

Francis, N., Ramey, V. A., Uhlig, H., and Basu, S. (2004). The source of historical economic fluctuations: An analysis using long-run restrictions [with comments]. In *NBER International Seminar on Macroeconomics*, volume 2004, pages 17–73. The University of Chicago Press Chicago, IL.

Fry, R., Pagan, A., et al. (2005). Some issues in using vars for macroeconometric research. *Centre for Applied Macroeconomic Analyses, CAMA Working Paper*, 18.

Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American Economic Review*, 89(1):249–271.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Guiso, L., Sapienza, P., and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, 128(3):403–421.

Hall, A. R. (2005). *Generalized method of moments*. Oxford university press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in us labour productivity. *Journal of Economic perspectives*, 15(4):117–128.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Inoue, A., Jin, L., and Pelletier, D. (2020). Local-linear estimation of time-varying-parameter garch models and associated risk measures. *Journal of Financial Econometrics*.

Inoue, A. and Kilian, L. (2016). Joint confidence sets for structural impulse responses. *Journal of Econometrics*, 192(2):421–432.

Inoue, A., Rossi, B., and Wang, Y. (2022). Local projections in unstable environments: How effective is fiscal policy?

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

Jordà, Ò., Singh, S. R., and Taylor, A. M. (2020). The long-run effects of monetary policy. Technical report, National Bureau of Economic Research.

Jordà, Ò. and Taylor, A. M. (2016). The time for austerity: estimating the average treatment effect of fiscal policy. *The Economic Journal*, 126(590):219–255.

Kalirajan, K. (1999). Stochastic varying coefficients gravity model: an application in trade analysis. *Journal of Applied Statistics*, 26(2):185–193.

Kapetanios, G., Pagan, A., and Scott, A. (2007). Making a match: Combining theory and evidence in policy-oriented macroeconomic modeling. *Journal of Econometrics*, 136(2):565–594.

Kascha, C. and Mertens, K. (2009). Business cycle analysis and varma models. *Journal of Economic Dynamics and Control*, 33(2):267–282.

Kascha, C. and Trenkler, C. (2015). Simple identification and specification of cointegrated varma models. *Journal of Applied Econometrics*, 30(4):675–702.

Kim, C.-J., Nelson, C. R., et al. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.

Kristensen, D. and Lee, Y. J. (2019). Local polynomial estimation of time-varying parameters in nonlinear models. *arXiv preprint arXiv:1904.05209*.

Li, D., Plagborg-Møller, M., and Wolf, C. K. (2022). Local projections vs. vars: Lessons from thousands of dgps. Technical report, National Bureau of Economic Research.

Lütkepohl, H. (2005). Estimation of varma models. In *New introduction to multiple time series analysis*, pages 447–492. Springer.

Lütkepohl, H. (2006). Forecasting with varma models. *Handbook of economic forecasting*, 1:287–325.

Marimoutou, V., Peguin, D., and Peguin-Feissolle, A. (2010). The" distance-varying" gravity model in international economics: is the distance an obstacle to trade?

Mertens, K. and Ravn, M. O. (2011). Technology-hours redux: Tax changes and the measurement of technology shocks. In *NBER International Seminar on Macroeconomics*, volume 7, pages 41–76. University of Chicago Press Chicago, IL.

Metaxoglou, K. and Smith, A. (2007). Maximum likelihood estimation of varma models using a state-space em algorithm. *Journal of Time Series Analysis*, 28(5):666–685.

Miranda-Agrippino, S. and Rey, H. (2020). Us monetary policy and the global financial cycle. *The Review of Economic Studies*, 87(6):2754–2776.

Miranda-Agrippino, S. and Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107.

Newey, W. K. and McFadden, D. (1994a). Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier.

Newey, W. K. and McFadden, D. (1994b). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Nishihat, M. and Otsu, T. (2020). Conditional gmm estimation for gravity models. *Economics Bulletin*, 40(2):1106–1111.

Orbe, S., Ferreira, E., and Rodríguez-póo, J. (2000). A nonparametric method to estimate time varying coefficients under seasonal constraints. *Journal of nonparametric statistics*, 12(6):779–806.

Orbe, S., Ferreira, E., and Rodriguez-Poo, J. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *journal of Econometrics*, 126(1):53–77.

Orbe, S., Ferreira, E., and Rodriguez-Poo, J. (2006). On the estimation and testing of time varying constraints in econometric models. *Statistica Sinica*, pages 1313–1333.

Plagborg-Møller, M. (2017). Essays in macroeconometrics: Estimation of smooth impulse response functions. Chapter 3 of Harvard University PhD thesis.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980.

Poskitt, D. S. (2016). Vector autoregressive moving average identification for macroeconomic modeling: A new methodology. *Journal of econometrics*, 192(2):468–484.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72:821–852.

Raghavan, M., Athanasopoulos, G., and Silvapulle, P. (2016). Canadian monetary policy analysis using a structural varma model. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(1):347–373.

Raghavan, M., Athanasopoulos, G., Silvapulle, P., et al. (2009). Varma models for malaysian monetary policy analysis. Technical report, Monash University, Department of Econometrics and Business Statistics.

Ramey, V. (2016). Macroeconomic Shocks and Their Propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*. Elsevier, Amsterdam, North Holland.

Ramey, V. A. (2011). Can government purchases stimulate the economy? *Journal of Economic Literature*, 49(3):673–85.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, 54(7):2048–2064.

Rey, H. (2015). Dilemma not trilemma: the global financial cycle and monetary policy independence. Technical report, National Bureau of Economic Research.

Richter, S. and Dahlhaus, R. (2019). Cross validation for locally stationary processes. *The Annals of Statistics*, 47(4):2145 – 2173.

Robinson, P. (1989). Nonparametric Estimation of Time-Varying Parameters. In Hackl, P., editor, *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer, Berlin.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change*, pages 179–190. Springer.

Romer, C. D. and Romer, D. H. (2004a). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94:1055–1084.

Romer, C. D. and Romer, D. H. (2004b). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084.

Ruisi, G. (2019). Time-varying local projections. Technical report, Working Paper.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, pages 1–48.

Sims, C. A. (1999). Drifts and breaks in monetary policy. Technical report, working paper, Yale University.

Sims, C. A. and Zha, T. (1999). Error bands for impulse responses. *Econometrica*, 67(5):1113–1155.

Sims, C. A. and Zha, T. (2001). Stability and instability in us monetary policy behavior. *Unpublished Paper*.

Sims, C. A. and Zha, T. (2006). Were there regime switches in u.s. monetary policy? *American Economic Review*, 96(1):54–81.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

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.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948.

Su, L., Murtazashvili, I., and Ullah, A. (2013). Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling. *Journal of Business & Economic Statistics*, 31(2):184–207.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–195.

Triantafyllopoulos, K. and Nason, G. P. (2007). A bayesian analysis of moving average processes with time-varying parameters. *Computational statistics & data analysis*, 52(2):1025–1046.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *journal of the american statistical association*, 93(443):1188–1202.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.

Tzouvelekas, V. (2007). Accounting for pairwise heterogeneity in bilateral trade flows: a stochastic varying coefficient gravity model. *Applied Economics Letters*, 14(12):927–930.

Uhlig, H. (2004). Do technology shocks lead to a fall in total hours worked? *Journal of the European Economic Association*, 2(2-3):361–371.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.

Yan, Y., Gao, J., and Peng, B. (2020). Nonparametric time-varying vector moving average (infinity) models. *Available at SSRN 3729872*.

Yao, W., Kam, T., and Vahid, F. (2017). On weak identification in structural varma models. *Economics letters*, 156:1–6.