Functional VAR*

Jiaming Huang

Universitat Pompeu Fabra and Barcelona School of Economics

May 12, 2024

Abstract

This paper models the joint dynamics of macro aggregates and functional variables within the Structural VAR framework. I reduce the dimension of the system using functional PCA and show that the proposed functional VAR (FVAR) consistently recovers the responses of the functions. The FVAR is easy to implement and fully compatible with conventional SVAR tools. Simulation evidence shows that it performs satisfactorily in finite samples. Applying FVAR to study the impact of tax shocks on income distributions in the UK, I find that tax cuts persistently reduce the density of lower-middle-class households, which is offset by a substantial increase in the richer range and a moderate increase in the poorer range. However, this pattern is not captured by VARs with conventional inequality measures.

Keywords: functional principal components, structural VAR, income distributions

JEL:C32, C38, D31, E32, E62

*Contact: jiaming.huang@upf.edu. I would like to thank Geert Mesters, Barbara Rossi, Christian Brownlees, Davide Debortoli, Jordi Galí, Priit Jeenas, Regis Barnichon, Weixuan Zhu and seminar participants at CREI Macro Lunch, UPF Econometrics Workshop, the 2020 Symposium on Modern Statistics in Xiamen, and the XIV_t Workshop in Time Series Econometrics in Zaragoza for helpful comments. I gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (FPI Grant PRE2020-092551). All errors are my own.

1 Introduction

What is the impact of economic policy on income distributions? Does inequality affect business cycles? Although the questions have received increasing attention in recent decades (Benhabib et al., 2011; Jappelli and Pistaferri, 2014; Colciago et al., 2019), very few attempt to empirically answer them directly. One major obstacle is that distributions are intrinsically infinite dimensional functions, which poses a great challenge to standard econometric methods.

Instead, the most popular method is to replace the distribution functions by summary statistics —for example percentiles, moments and Gini coefficients— and proceed with small-scale VARs or the local projection counterparts (e.g. Coibion et al., 2017; **?**). However, I show in the empirical application that policy implications change if we look at the entire income distributions instead of the inequality measures.

This paper proposes a general framework to study the joint dynamics of macro aggregates and the *entire* distributions building on recent advances in functional analysis (Horváth and Kokoszka, 2012).[1] Specifically, I augment a standard VAR model with the distribution functions and reduce the dimension of the system though functional PCA. The model therefore consists of two main steps. First, researchers extract the functional principal component scores, which capture the dynamics of the underlying functions. Second, substitute the FPC scores for the functions and estimate the resulting VAR by standard techniques; see Kilian and Lütkepohl (2017) for a review. In particular, I show that the functional VAR procedure is able to consistently recover the impulse responses of the distribution functions.

A simulation study is conducted to evaluate the functional VAR framework in finite samples. I simulate data from a functional VAR model and estimate the functional responses using SVAR with internal instrument (Plagborg-Møller and Wolf, 2021). Simulation evidence shows that the estimation errors are moderate and the coverage rates are satisfactory when the underlying true responses are regular, i.e. hump-shaped. In contrast, it fails to generate

[1] The framework is applicable to general functions, e.g. yield curves.

responses of irregular shapes.

Within the functional VAR framework I revisit the impact of tax policy on income distributions in the United Kingdom from 1968Q1 to 2009Q4. Specifically, I estimate quarterly household income distributions using detailed data from the Households Below Average Income Dataset (HBAI).[2] I then estimate the functional VAR model with exogenous tax changes constructed by Cloyne (2013). I find that following a tax cut, there is a sharp and persistent decrease in the densities of households with weekly income between £100 − £300. Moreover, such decrease is compensated for by large increases in the densities of the households with income over £300, and only a moderate increase in the extremely poor range. Furthermore, counterfactual analysis shows that tax policy explains a non-negligible fraction of the cyclical fluctuations in the income distributions.

The findings reflect two opposing forces. On the one hand, tax cuts boost the economy and households benefit from higher employment and income (Cloyne, 2013; Ljungqvist and Smolyansky, 2016). On the other hand, tax cuts are generally associated with spending cut especially during fiscal consolidation (Erceg and Lindé, 2013; Glomm et al., 2018), which may hurt the poor. Our results indicate that the former effect dominates, but tax cuts should also be complemented by targeted policy to the extremely poor. Importantly, the observed pattern is also consistent with the literature that tax cuts unambiguously increases conventional inequality measures, e.g. Gini coefficients (Clark and Leicester, 2005; Coady and Gupta, 2012). Nonetheless, given the responses of the entire distribution, the rising inequality should be interpreted as a natural outcome of the shrinking middle class, instead of an alarming situation that cautions against tax cuts.

Alternative methods to study functional variables have been proposed. Chang et al. (2022) use sieve approximations to decompose the functional variables. Our framework differs in two important aspects. First, the basis functions in our functional VAR are eigenfunctions of the data covariance function, and are optimal in the sense that it maximizes the variations

[2] The HBAI dataset contains income data for *all* surveyed households, excluding the short-term self-employed and temporarily separated couples.

explained given the same number of basis. Second, our approach can be estimated easily by OLS whereas the state-space model in Chang et al. (2022) is estimated by Bayesian estimation.[3] Another related method is the "*VARs with functional shocks*" by Inoue and Rossi (2021) where they rely on certain parametric models to decompose functional variables. In contrast, our framework is non-parametric and can handle general functions in the L^2 space.

Our approach also connects to the burgeoning literature on functional linear regressions and functional autoregressions (Bosq, 2000; Shin, 2009; Kokoszka and Reimherr, 2013; Hörmann and Kidziński, 2014). However, in the current setup, I establish asymptotic properties for mixed-type data that contains errors from density estimation and exhibits weak dependence. Moreover, the object of interest is the coefficient (and thus the impulse responses), instead of the prediction (Cai and Hall, 2006; Aue et al., 2015).

The remainder of the paper is organized as follows. Section 2 illustrates the functional VAR framework through a simple example. Section 3 formalizes the model and provides details on the estimation. Section 4 discusses the asymptotic properties and Section 5 evaluates the finite sample performance. Section 6 presents the empirical application and Section 7 concludes.

2 An illustrative example

Suppose we are interested in the impact of TFP shocks on household income distributions, we can model it by

$$\begin{array}{rcl} y_t & = & a_{yy}y_{t-1} + \int a_{yf}(u)f_{t-1}(u)du + e_{yt} \\ f_t(u) & = & a_{fy}(u)y_{t-1} + \int a_{ff}(u,v)f_{t-1}(v)dv + e_{ft}(u) \end{array}, \quad t = 1, \ldots, T \tag{1}$$

[3] Another recent work within the Bayesian paradigm is Huber et al. (2023) where the dynamics of functional scores and the aggregates are modeled by a Bayesian VAR.

where y_t is the GDP growth and $f_t(u)$ is the income density function. For illustrative purpose, I assume that both y_t and $f_t(u)$ have zero mean.

Although the proposed formulation is appealing, direct estimation is challenging because the model is of infinite dimension. One may therefore propose to replace f_t by its realizations on some finite grid $[f_t(u_1), \ldots, f_t(u_N)]$. This approach however faces a sharp bias-variance trade-off: A large grid ensures accurate approximation but inflates estimation errors, whereas a small grid discards valuable information and results in omitted variable bias. Alternatively, one may prefer summary statistics of f_t, e.g. Gini coefficients. However, such measures are not sufficient statistics and important dynamics of the functions could be missed.

In contrast, the functional VAR procedure balances well the bias-variance trade-off. The main idea is that with functional PCA techniques, we can decompose f_t into eigenfunctions and *finite dimensional* functional principal component (FPC) scores. Since FPC scores summarize how the functions fluctuate, we can substitute them for the functions and estimate the resulting VAR with standard methods. Specifically, the functional PCA procedure is summarized as follows.

Functional VAR

1. *Density Estimation.* Obtain consistent estimates of density functions, denoted by $\hat{f}_t$.

2. *FPCA.* Decompose the functions by functional principal component analysis, with truncation order q

$$\hat{f}_t(u) = \sum_{j=1}^{q} \hat{\eta}_{jt} \hat{\xi}_j(u) \tag{2}$$

   where $\hat{\eta}_{jt}$ is the estimated FPC scores and $\hat{\xi}_j$ the eigenfunctions.

3. *SVAR* Substitute the FPC scores for functions, leading to

$$\begin{bmatrix} y_t \\ \hat{\eta}_t \end{bmatrix} = \begin{bmatrix} a_{yy} & a_{y\eta} \\ a_{\eta y} & a_{\eta\eta} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \hat{\eta}_{t-1} \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{\eta t} \end{bmatrix}. \tag{3}$$

where $\hat{\eta}_t = [\hat{\eta}_{1,t}, \ldots, \hat{\eta}_{q,t}]'$ is a $q \times 1$ vector. Denote the estimated impulse responses of $\hat{\eta}_{jt}$ by $\hat{\theta}_{h,j}$, then the responses of f_t can be estimated by $\hat{\Theta}_h = \sum_{j=1}^{q} \hat{\theta}_{h,j}\hat{\xi}_j$.

The functional VAR procedure bears remarkable similarity to the two step estimation of FAVAR in Bernanke et al. (2005). In particular, FAVAR characterizes a high dimensional data matrix through a few factors, whereas FVAR summarizes the infinite dimensional functions by a few FPC scores. Further, in the same way as FAVAR recovering the responses of information variables by factor responses and static factor loadings, the FVAR estimates functional responses by multiplying the responses of FPC scores and the static eigenfunctions. As a whole, our approach extends FAVAR to the case when the cross-section dimension of the data *is* infinite.

3 FVAR model and estimation

I now formalize the functional VAR model and provide details for estimation.

3.1 FVAR model

Assume that we observe data $\{(Y_1, f_1), \ldots, (Y_T, f_T)\}$ where Y_t is a $K \times 1$ vector of aggregates and f_t is a 1×1 function on a separable Hilbert space. Consider a general FVAR(p) model

$$\begin{aligned} Y_t &= \textstyle\sum_{i=1}^{p} \Psi_{11,i} Y_{t-i} + \sum_{i=1}^{p} \Psi_{12,i}(f_{t-i}) + e_{yt} \\ f_t(u) &= \textstyle\sum_{i=1}^{p} \Psi_{21,i}(u) Y_{t-i} + \sum_{i=1}^{p} \Psi_{22,i}(f_{t-1})(u) + e_{ft}(u) \end{aligned} \tag{4}$$

where $\Psi_{11,i}$ is a $K \times K$ matrix, $\Psi_{21,i} = (\Psi_{21,1}, \ldots, \Psi_{21,K})$ is a $1 \times K$ vector of functions, and

$$\begin{aligned} \Psi_{12,i}(f_{t-i}) &= \textstyle\int \psi_{12,i}(u) f_{t-i}(u) du \\ \Psi_{22,i}(f_{t-i})(u) &= \textstyle\int \psi_{22,i}(u,v) f_{t-i}(v) dv \end{aligned} . \tag{5}$$

For ease of exposition, I assume that both Y_t and f_t have zero means and normalize the domain u to $[0, 1]$.

Model (4) consists of two blocks: The first block is the scalar-on-function regressions (Cardot et al., 1999; Hall and Horowitz, 2007; Shin, 2009) and the second falls into the category of functional autoregressions (Bosq, 2000; Kokoszka and Reimherr, 2013). This motivates the use of principal component regression methods (Ramsay and Silverman, 2005).

Specifically, define the covariance function of f_t as $\gamma_f(u,v) = E[f(u)f(v)]$ and the covariance operator Γ_f is $\Gamma_f(g) = \int \gamma_f(u,v)g(v)dv$. The eigenfunctions $\{\xi_j\}_{j=1}^{\infty}$ and the associated eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$ are determined by

$$\Gamma_f(\xi_j)(u) = \int \gamma_f(u,v)\xi_j(v)dv = \lambda_j \xi_j(u). \tag{6}$$

Assume without loss of generality that the eigenfunctions are normalized to unit norm. Then eigenfunctions $\{\xi_j\}_{j=1}^{\infty}$ form an orthonormal basis in the functional space. We have

$$f_t(u) = \sum_{j=1}^{\infty} \eta_{t,j}\xi_j(u) \tag{7}$$

where the FPC scores $\eta_{t,j}$ are given by $\eta_{t,j} = \int f_t(u)\xi_j(u)du$. Importantly, the above decomposition is optimal in the sense that for all orthonormal basis $\{\nu_1, \nu_2, \ldots, \nu_q\}$,

$$\sum_{t=1}^{T} \left\| f_t(u) - \sum_{j=1}^{q} \eta_{t,j}\nu_j(u) \right\|^2$$

is minimized at $\{\xi_1, \ldots, \xi_q\}$ for all q.

As in Section 2, applying the functional PCA to model (4) leads to

$$\begin{aligned} Y_t &= \textstyle\sum_{i=1}^{p} \Phi_{11,i} Y_{t-i} + \sum_{i=1}^{p}\sum_{j=1}^{\infty} \Phi_{12,i,j}\eta_{t-i,j} + e_{yt} \\ \textstyle\sum_{j=1}^{\infty} \eta_{t,j}\xi_j &= \textstyle\sum_{i=1}^{p}\sum_{j=1}^{\infty} \Phi_{21,i,j}\xi_j Y_{t-i} \\ &\quad + \textstyle\sum_{i=1}^{p}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \Phi_{22,i,jk}\eta_{t-i,j}\xi_j + \sum_{j=1}^{\infty} e_{\eta t,j}\xi_j \end{aligned} \tag{8}$$

which can be rewritten as

$$\begin{bmatrix} 1 & 0 \\ 0 & \xi_1 \\ 0 & \xi_2 \\ \dots & \dots \end{bmatrix}' \begin{bmatrix} Y_t \\ \eta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \xi_1 \\ 0 & \xi_2 \\ \dots & \dots \end{bmatrix}' \left(\begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} Y_t \\ \eta_t \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{\eta t} \end{bmatrix} \right) \tag{9}$$

where $\eta_t = [\eta_{t,1}, \eta_{t,2}, \dots]'$ is a vector of FPC scores and $\Psi_{\cdot,\cdot}(L)$ are coefficient matrix, both of which are infinite dimensional. Clearly, the population model cannot be estimated directly. Instead, we truncate (7) at some order q, which leads to the truncated score VAR

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \xi_1 & \dots & \xi_q \end{bmatrix} [I - \Psi(L)] \underbrace{\begin{bmatrix} Y_t \\ \eta_t \end{bmatrix}}_{(K+q)\times 1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \xi_1 & \dots & \xi_q \end{bmatrix} \begin{bmatrix} e_{yt} \\ e_{\eta t} \end{bmatrix}. \tag{10}$$

Finally, I discuss briefly the identification of FVAR. Notice from (10) that

$$\begin{bmatrix} e_{yt} \\ \sum_{j=1}^{q} e_{\eta t,j}\xi_j \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \xi_1 \\ \dots & \\ 0 & \xi_q \end{bmatrix}' \begin{bmatrix} e_{yt} \\ e_{\eta t,1} \\ \dots \\ e_{\eta t,q} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \xi_1 \\ \dots & \\ 0 & \xi_q \end{bmatrix}' \underbrace{B}_{(K+q)\times(K+q)} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{1t} \\ \dots \\ \epsilon_{qt} \end{bmatrix} \tag{11}$$

where $\epsilon_{\cdot,t}$ are structural shocks. Analogous to standard VAR, matrix B is the object to be identified. Equations (10) and (11) demonstrate clearly that to construct the impulse responses, we need only to focus on the score VAR. Suppose we want to recover the responses of f_t to the TFP shock ϵ_{yt}. Denote the first column of B as $B_1 = [b_{yy}, b_{y\eta}]'$, the response of $\eta_{t,j}$ to TFP shocks at horizon h is

$$\theta_{h,j} \triangleq \frac{\partial \eta_{t+h,j}}{\partial \epsilon_{yt}} = [I - \Psi(L)]^h b_{y\eta,j} \tag{12}$$

and the response of f_t is

$$\Theta_h \triangleq \frac{\partial f_{t+h}}{\partial \epsilon_{yt}} = \frac{\partial \left(\sum_{j=1}^{\infty} \eta_{t+h,j} \xi_j \right)}{\partial \epsilon_{yt}} = \sum_{j=1}^{\infty} \frac{\partial \eta_{t+h,j}}{\partial \epsilon_{yt}} \xi_j \approx \sum_{j=1}^{q} \theta_{h,j} \xi_j \tag{13}$$

which can be estimated by plug in the estimates $\hat{A}$, $\hat{B}$ and $\hat{\xi}_j$.

Although the current paper focuses on the general FVAR framework and leaves the identification of functional shocks for future research, $\{\epsilon_{jt}\}_{j=1}^{q}$ can be easily interpreted when the eigenfunctions have structural interpretations (e.g., Benko, 2007).

3.2 Estimation

The formulations in the previous section assumes that both functions f_t and the principal components are perfectly observed. In practice, these objects need to be estimated.

First, we estimate the densities from the data. Note that f_t is often defined on a bounded support, e.g. salaries cannot be negative, rendering standard kernel density estimation inconsistent. One solution is to use the modified kernel density estimation proposed by Petersen and Müller (2016). Specifically, let κ be a kernel that corresponds to a continuous probability density function and $h < 1/2$ be the bandwidth. We can estimate the density from an i.i.d. sample $\{x_1, \ldots, x_N\}$ by

$$\hat{f}(u) = \sum_{l=1}^{N} \kappa \left(\frac{u - x_l}{h} \right) w(u,h) \Bigg/ \sum_{l=1}^{N} \int_0^1 \kappa \left(\frac{v - x_l}{h} \right) w(v,h) dv \tag{14}$$

for $u \in [0,1]$ and 0 elsewhere. The weight function is designed to remove the boundary bias:

$$w(u,h) = \begin{cases} \left(\int_{-u/h}^{1} \kappa(v) dv \right)^{-1} & u \in [0,h) \\ \left(\int_{-1}^{(1-u)/h} \kappa(v) dv \right)^{-1} & u \in (1-h,1] \\ 1 & o.w. \end{cases} . \tag{15}$$

Second, to conduct functional PCA we estimate the covariance functions by the sample analogue

$$\hat{\gamma}(u,v) = \frac{1}{T}\sum_{t=1}^{T} \hat{f}_t(u)\hat{f}_t(v). \tag{16}$$

The eigenfunction $\xi_j(u)$ is obtained by solving the eigen-equation

$$\int \hat{\gamma}(u,v)\hat{\xi}_j(v)dv = \hat{\lambda}_j\hat{\xi}_j(u)\ . \tag{17}$$

Since the functions are infinite dimensional, the eigen-equation is in fact solved by discretizing the covariance functions (Rice and Silverman, 1991). Given $\hat{\xi}_j$, the FPC scores are simply the generalized Fourier coefficients when projecting $\hat{f}_t$ on the eigenfunctions. Given the pair of eigen-elements $(\hat{\lambda}_j, \hat{\xi}_j)$ we need to truncate it at some order q. Several methods have been proposed: For instance, it can be determined by the fraction of variation explained (Ramsay and Silverman, 2005), leave-one-out cross-validation (Rice and Silverman, 1991) or the information criterion (Yao et al., 2005; Li et al., 2013). Implementation of the above steps is facilitated through the Functional Data Analysis (FDA) package.[4]

4 Asymptotic theory

This section establishes asymptotic properties of the functional VAR procedure. We start by introducing a set of useful notations in Section 4.1. Section 4.2 discusses consistency of the FVAR procedure . Proofs are given in Appendix A.

4.1 Notation

Denote by L^2 the space of square integrable functions such that $\int f(u)^2 du < \infty$ for all $f \in L^2$. The space is associated with the inner product $\langle f, g\rangle = \int f(u)g(u)du$ for $f, g \in L^2$, and the norm $\|f\| = \sqrt{\langle f, f\rangle} = \sqrt{\int f^2(u)du}$. Moreover, denote by $\mathcal{L}$ the space of continuous

[4]The package is provided by Ramsay and Silverman (2005) and can be accessed here. An alternative package PACE based on Yao et al. (2005) yields almost identical results.

linear operators from L^2 to L^2. For all $\Psi \in \mathcal{L}$, the uniform norm of the operators is $\|\Psi\|_{op} = \sup\{\|\Psi(g)\| : g \in L^2 \text{ with } \|g\| \leq 1\}$, and the Hilbert-Schmidt norm is $\|\Psi\|_{\mathcal{HS}} = (\sum_{j=1}^{\infty} \|\Psi(\nu_j)\|^2)^{1/2}$ where $\{\nu_j\}_{j=1}^{\infty}$ is any orthonormal basis in L^2. Finally, denote $f \circ g : \langle f, \cdot \rangle g$. Then the covariance operator can be written as $\Gamma_f = E[\langle f, \cdot \rangle f] = E[f \circ f]$.

Next, assume that all random variables are defined on some common probability space $(\Omega, \mathcal{A}, P)$. A function f has p moments if $(E\|f\|_{L2}^p)^{1/p} < \infty$. Furthermore, we call a sequence of $\epsilon_t \in L^2$ an H-white noise if it has zero mean $E\epsilon_t = 0$, constant covariance $\Gamma_{\epsilon_t} = \Gamma_{\epsilon_0}$, and are mutually uncorrelated $\epsilon_s \perp \epsilon_t,\ \forall s \neq t$.

Moreover, I collect the lags of model (4) as in Kokoszka and Reimherr (2013). Define $I_i(u)$ an indicator function that equals 1 if $u \in ((i-1)/p, i/p]$ and

$$\begin{array}{lcl} W_t(u) & = & \sum_{i=1}^{p} f_{t-i}(up - (i-1)) I_i(u) \\ \psi_{12}(u) & = & \sum_{i=1}^{p} p\psi_{12,i}(up - (i-1)) I_i(u) \\ \psi_{22}(u,v) & = & \sum_{i=1}^{p} p\psi_{22,i}(u, vp - (i-1)) I_i(v) \end{array} \tag{18}$$

$$\begin{aligned} \Psi_{12}(W_t) &= \sum_{i=1}^{p} \Psi_{12,i}(f_{t-i}) \\ &= \int_0^1 \sum_{i=1}^{p} pI_i(r)\psi_{12,i}(rp-(i-1)) f_{t-i}(rp-(i-1))dr \\ \Psi_{22}(W_t) &= \sum_{i=1}^{p} \Psi_{22,i}(f_{t-i}) \\ &= \int_0^1 \sum_{i=1}^{p} pI_i(r)\psi_{22,i}(u, rp-(i-1)) f_{t-i}(rp-(i-1))dr \end{aligned} . \tag{19}$$

Denote $Z_t = (Y_{t-1}', \ldots, Y_{t-p}')'$, $\Psi_{11} = (\Psi_{11,1}, \ldots, \Psi_{11,p})$ and $\Psi_{21} = (\Psi_{21,1}, \ldots, \Psi_{21,p})$. The dimensions are $Kp \times 1$, $K \times Kp$ and $1 \times Kp$ respectively. Then model (4) can be written compactly as

$$\begin{array}{lcl} Y_t & = & \Psi_{11} Z_t + \Psi_{12}(W_t) + e_{yt} \\ f_t & = & \Psi_{21} Z_t + \Psi_{22}(W_t) + e_{ft} \end{array} \tag{20}$$

and in the matrix form

$$\begin{array}{lll} \underbrace{Y}_{K\times T} & = & \Psi_{11}\underbrace{Z}_{Kp\times 1}+\Psi_{12}(\underbrace{W}_{1\times T})+e_y \\ f & = & \Psi_{21}Z+\Psi_{22}(W)+e_f \end{array}. \tag{21}$$

The score VAR can be written similarly. To derive it, first denote the covariance operator of W_t as Γ_W. We have eigen-equation $\Gamma_W(\phi_j)=\rho_j\phi_j$ where ρ_j is the eigenvalues and ϕ_j is the eigenfunctions. Functional PCA on W_t gives the decomposition

$$W_t(u)=\sum_{j=1}^{\infty}\pi_{t,j}\phi_j(u) \tag{22}$$

where π_j are the FPC scores. Truncate (7) and (22) at order q_f and q_W respectively.[5] Denote $\eta_t=(\eta_{t,1},\dots,\eta_{t,q_f})'$ and $\pi_t=(\pi_{t,1},\dots,\pi_{t,q_W})'$. We have the score VAR[6]

$$\begin{array}{lll} Y_t & = & \Psi_{11}Z_t+\Psi_{12}\hat{\pi}_t+\check{e}_{yt} \\ \hat{\eta}_t & = & \Psi_{21}Z_t+\Psi_{22}\hat{\pi}_t+\check{e}_{\eta t} \end{array} \tag{23}$$

which in matrix form, is

$$\begin{array}{lll} Y & = & \Psi_{11}Z+\Psi_{12}\hat{\pi}+\check{e}_y \\ \hat{\eta} & = & \Psi_{21}Z+\Psi_{22}\hat{\pi}+\check{e}_\eta \end{array}. \tag{24}$$

Here, I use $\check{e}_y$ and $\check{e}_\eta$ to explicitly indicates the additional errors due to truncation.

Applying the Frisch-Waugh-Lovell theorem, the functional VAR estimators are

$$\begin{array}{ll} \hat{\Psi}_{11}=Y\hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1}, & \hat{\Psi}_{12}=YM_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1} \\ \hat{\Psi}_{21}=\hat{\eta}\hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1}, & \hat{\Psi}_{22}=\hat{\eta}M_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1} \end{array} \tag{25}$$

where $\hat{M}_\pi=I-\hat{\pi}'(\hat{\pi}\hat{\pi}')^{-1}\hat{\pi}$ and $M_Z=I-Z'(ZZ')^{-1}Z$ are $T\times T$ residual makers. These are

[5] Note that we treat W_t and f_t separately simply for ease of notations. In practice, we will use the p lags of $\hat{f}_t$ instead of $\hat{W}_t$. Therefore, we have $q_W=q_f*p$.

[6] With a bit abuse of notation, Ψ_{21} is a $1\times Kp$ vector of functions in (20) whereas it stands for a $q_f\times Kp$ matrix in the score VAR.

the celebrated principal component estimators in the functional linear regression literature (Cardot et al., 1999; Shin, 2009; Kokoszka and Reimherr, 2013). Moreover, we have

$$\begin{aligned}\hat{\psi}_{12}(u) &= \textstyle\sum_{j=1}^{q_W} \hat{\Psi}_{12}[\cdot, j]\hat{\phi}_j(u), \quad \hat{\psi}_{21}(u) = \sum_{j=1}^{q_f} \hat{\Psi}_{21}[j, \cdot]\hat{\xi}_j(u) \\ \hat{\psi}_{22}(u, v) &= \textstyle\sum_{j=1}^{q_f} \sum_{k=1}^{q_W} \hat{\Psi}_{22}[j, k]\hat{\xi}_j(u)\hat{\phi}_k(v)\end{aligned} \tag{26}$$

where $\hat{\Psi}_{12}[\cdot, j]$ is the *j-th* column of $\hat{\Psi}_{12}$, $\hat{\Psi}_{21}[j, \cdot]$ is the *j-th* row of $\hat{\Psi}_{21}$ and $\hat{\Psi}_{22}[j, k]$ is the (j, k) element of $\hat{\Psi}_{22}$. We can easily recover operators in model (20) by (26). For instance, $\hat{\Psi}_{12}(W_t) = \int \sum_{j=1}^{q_W} \hat{\Psi}_{12}[\cdot, j]\hat{\phi}_j(u)W_t(u)du$, which gives a $K \times 1$ vector.

4.2 Consistency

We are now ready to state the assumptions for consistency. Note that by construction, assumptions made on f_t applies to W_t. Specifically, I postulate that:

Assumption 1 (L^4-m-approximable)**.**

(i) The sequence $\{f_t\} \in L^2$ *satisfies* $E\|f\|^4 < \infty$.

(ii) f_t *is generated by* $f_t = h(v_t, v_{t-1}, \dots)$ *where* v_t *are i.i.d. error functions in some measurable space* S*, and* h *is some measurable function* $h : S^\infty \mapsto H$.

(iii) Let $f_t^{(m)} = h(v_t, v_{t-1}, \dots, v_{t-m+1}, v_{t-m}^{(t)}, v_{t-m-1}^{(t)}, \dots)$ *where for each* t*,* $\{v_s^{(t)}\}$ *is an independent copy of* $\{v_s\}$*. We have*

$$\sum_{m=1}^{\infty} \left(E\|f_m - f_m^{(m)}\|^4\right)^{\frac{1}{4}} < \infty. \tag{27}$$

Assumption 1 is introduced by Hörmann and Kokoszka (2010) which allows the density functions to exhibit weak dependence. The idea is that for each $\{f_t\}$ we can find an auxiliary sequence $\{f_t^{(m)}\}$ that is by construction m-dependent. Then $\{f_t\}$ will also be m-dependent as it converges to $\{f_t^{(m)}\}$. In this regard, the approximability assumption implies m-dependence and thus stationarity.

Specifically, it contains three parts. First, part 1(i) requires the functions to have finite fourth moments, which is used to bound auto-correlations of L^2 functions. In contrast, finite second moments suffice to guarantee consistency for i.i.d. functions (see Hall and Hosseini-Nasab, 2006). Second, 1(ii) presents the structural form of the random functions: It assumes that the f_t is generated by a general (possibly nonlinear) function of i.i.d. error functions. Hence it includes a broader class of random functions than linear processes discussed in Bosq (2000). Finally, 1(iii) introduces the auxiliary sequence, which is constructed by replacing the noises $\{\epsilon_{t-j}\}_{j\geq m}$ with independent copies $\{\epsilon_{t-j}^{(t)}\}_{j\geq m}$. The superscript (t) indicates that the copies are drawn independently for each period t, implying that $\{f_t^{(m)}\}$ is m-dependent and strictly stationary. Finally, equation (27) states that $\{f_t\}$ converges to the auxiliary sequence and thus inherits dependency.

The next assumption imposes regularity conditions on the density functions and the kernel used for estimation. Intuitively, it guarantees that the modified kernel density estimator (14) is consistent.

Assumption 2 (Density Estimation)**.**

(i) *For all $f \in \mathcal{F} \subset L^2$, f is differentiable. Moreover, there is a positive constant M such that $M \geq \max\{\|f\|_\infty, \|1/f\|_\infty, \|f'\|_\infty\}$ for all $f \in \mathcal{F}$.*

(ii) *The kernel κ is of bounded variation and is symmetric about 0. Moreover, we have $\int_0^1 \kappa(u)du > 0$, and $\int_{\mathbb{R}} |u|\kappa(u)du$, $\int_{\mathbb{R}} \kappa^2(u)du$ and $\int_{\mathbb{R}} |u|\kappa^2(u)du$ are finite.*

(iii) *For each period t, we have a random sample $\{x_1, \ldots, x_{N(t)}\} \overset{i.i.d.}{\sim} f_t$.*

The assumption contains three parts. First, 2(i) assumed that the density functions are continuously differentiable. Moreover, the density, its inverse and the first derivative are all bounded, requiring that $f(u) > 0$ for all $u \in [0, 1]$. Second, 2(ii) includes a set of regularity conditions on the kernel κ. In practice, common kernels such as the standard normal density suffices. Finally, it is assumed that for each period we obtain a random sample drawn from the density function. Importantly, the condition is satisfied even though the densities across time are dependent.

With the above assumptions, we can establish the consistency of functional PCA.

Proposition 1 (Consistent FPCA)**.** Under Assumptions 1-2, we have

$$\begin{aligned} \|\mu-\hat{\mu}\| &= O_p\left(T^{-1/2}\right), \quad \|\gamma_f-\hat{\gamma}_f\| = O_p\left(T^{-1/2}\right) \\ \left|\lambda_k-\hat{\lambda}_k\right| &= O_p\left(T^{-1/2}\right), \quad \|\xi_k-\hat{\xi}_k\| = O_p\left(T^{-1/2}\right) \end{aligned} \tag{28}$$

Two comments are in order. First, Proposition 1 states that even though the functions are estimated, we still achieve root-T consistency comparable to Hörmann and Kokoszka (2010). Importantly, as is shown in the proof, the results hold if the convergence of density estimation is fast enough, i.e. $O(h+(Nh)^{-1})$ to be finite.

Second, the consistency holds for *every* k. Intuitively, the result suggests that if the truncated (8) is the true model, then it is easy to show that the functional coefficients are consistent. To bridge the gap between the truncated, finite dimensional VAR and the population model, we introduce a final set of assumptions.

Assumption 3 (Truncation)**.**

(i) The eigenvalues λ_i are mutually distinct and in decreasing order $\lambda_1 > \lambda_2 > \cdots > 0$.

(ii) Denote $\alpha_i = \min\{\lambda_{i-1}-\lambda_i, \lambda_i-\lambda_{i+1}\}$ for $i \geq 2$ and $\alpha_1 = \lambda_1-\lambda_2$. Further define $R_1 = \operatorname{argmax}\{j \geq 1 \mid \hat{\lambda}_j \geq m_T^{-1}\}$ and $R_2 = \operatorname{argmax}\{k \geq 1 \mid \max_{1\leq j\leq k}\hat{\alpha}_j \geq m_T^{-1}\}$ where $m_T \to \infty$ such that $m_T^6 = o(T)$. The truncation order is set to be

$$q = \min\{R_1, R_2, m_T\}. \tag{29}$$

Assumption 3 involves two conditions. The first part is a standard identification assumption in the functional analysis literature (Cai and Hall, 2006; Hall and Horowitz, 2007). In particular, we require that $\lambda_1 > \lambda_2 > \cdots > 0$. The second condition introduced by Hörmann and Kidziński (2014) illustrates how the functional PCA balances the bias variance trade-off. Specifically, note that the eigenvalues measure the amount of variance explained by the associated eigenfunctions. Hence an additional FPC is included only when it explains a non-trivial portion of the functions, e.g. $\hat{\lambda}_j \geq m_T^{-1}$. Furthermore, it implies that the truncation order tends to infinity as $T \to \infty$.

Assumptions 1-3 guarantee that the functional PCA leads to accurate approximations of the functions, and the projection errors converges to zero as the sample size goes to infinity. However, so far we are silent about the regression itself. The following assumption imposes regularity conditions on the functional VAR model.

Assumption 4 (Functional VAR)**.**

(i) Y_t *is a stationary process and* e_{yt} *is iid white noise with nonsingular covariance matrix* Σ_{ey} *such that* $e_{yt} \overset{iid}{\sim} (0, \Sigma_{ey})$.

(ii) ψ_{12} *and* Ψ_{21} *are square integrable functions, and* Ψ_{22} *is a Hilbert-Schmidt operator. Further, the functional errors* e_{ft} *is an H-white noise.*

The above assumptions are fairly standard the literature. Specifically, part 4(i)is comparable to the assumptions in Kilian and Lütkepohl (2017, Chapter 2). An alternative way to formulate the stationarity condition (together with Assumption 1) is to impose restrictions on the coefficients (Kokoszka and Reimherr, 2013). The second part 4(ii) is standard in the functional regression literature. In particular, requiring the operators to be Hilbert-Schmidt, or assuming that the kernels reside in the L^2 space guarantees that the operators are bounded. Combined with Assumption 3(ii), the assumption ensures that the approximation errors shrink to zero.

We are now ready to state the consistency results.

Theorem 1 (Consistent FVAR)**.** Under Assumptions 1-4, as T goes to infinity we have

$$\begin{array}{lcl} \text{vec}(\hat{\Psi}_{11}) - \text{vec}(\Psi_{11}) & \xrightarrow{p} & 0, \quad \|\hat{\Psi}_{12} - \Psi_{12}\|_{op} \xrightarrow{p} 0 \\ \|\hat{\Psi}_{21} - \Psi_{21}\| & \xrightarrow{p} & 0, \quad \|\hat{\Psi}_{22} - \Psi_{22}\|_{op} \xrightarrow{p} 0 \end{array} \tag{30}$$

As is clear in Section 3, the impulse responses estimates are obtained by plugging in the estimates for $\Psi(L)$. Hence the consistency of the responses is implied by Theorem 1.

5 Simulation study

I conduct several simulation experiments to evaluate the finite sample performance of the functional VAR procedure. The DGP used is model (1) where the reduced-form errors are related to structural shocks through

$$\begin{array}{lcl} e_{yt} & = & \sigma_{11}\epsilon_{1t} + \int \sigma_{12}(u)\epsilon_{2t}(u)du \\ e_{ft}(u) & = & \sigma_{21}(u)\epsilon_{1t} + \int \sigma_{22}(u,v)\epsilon_{2t}(v)dv \end{array} \tag{31}$$

which, rewritten as combinations of the basis functions, is equivalent to

$$\begin{array}{lcl} e_{yt} & = & \sigma_{11}\epsilon_{1t} + \sigma_{12}[\int \xi'(u)\xi(u)du]\epsilon_{2t} \\ e_{ft}(u) & = & \xi(u)\sigma_{21}'\epsilon_{1t} + \xi(u)\sigma_{22}[\int \xi(v)'\xi(v)dv]\epsilon_{2t} \end{array}. \tag{32}$$

Here, σ_{12} and σ_{21} are $1 \times q$ vectors of scalar and σ_{22} is a $q \times q$ matrix, and $\epsilon_{1t}, \epsilon_{2t}$ are scalar structural shocks. Similar to what we have shown in Section 2, substitute the above decomposition into the FVAR(1) model, we have

$$\begin{array}{lcl} y_{t+1} & = & a_{yy}y_t + a_{yf}\eta_t + \sigma_{11}\epsilon_{1t} + \sigma_{12}\epsilon_{2t} \\ \xi(u)\eta_{t+1} & = & \xi(u)a_{fy}'y_t + \xi(u)a_{ff}\eta_t + \xi(u)\sigma_{21}'\epsilon_{1t} + \xi(u)\sigma_{22}\epsilon_{2t} \end{array}. \tag{33}$$

The scalar shocks ϵ_{1t} and ϵ_{2t} are drawn independently from standard normal distributions. Moreover, I generate instruments for ϵ_{1t} by $z_t = \epsilon_{1t} + \nu_t$ where $\nu_t \sim \mathcal{N}(0,1)$ are i.i.d noises.

The model is estimated using the internal instrument methods (Plagborg-Møller and Wolf, 2021). Specifically, the instrument is ordered first and the VAR is identified by Cholesky decomposition. Finally, the responses are computed for $H = 12$ horizons.

I employ two set of parameters as summarized in Table 1, with the associated impulse responses in Figure 1. As is clear, *Design 1* yields standard "hump-shaped" IRs whereas *Design 2* yields more irregular responses. For each parameter set, I simulate the data for 2000 runs and $T = 100, 200, 500$ respectively. Moreover, the basis function $\xi(u)$ is set to be Fourier basis functions on $[0, 1]$ of order three (q=3).

To evaluate how precise the FVAR procedure is, I report the mean integrated squared errors (MISE) and uniform errors (UE) of the IRs for all horizons. Specifically, the MISE is defined as

$$\text{MISE}(h) = \int_{u\in[0,1]} (\Theta_h(u) - \hat{\Theta}_h(u))^2 du \tag{34}$$

and the uniform errors

$$\text{UE}(h) = \sup_{u\in[0,1]} |\Theta_h(u) - \hat{\Theta}_h(u)|. \tag{35}$$

Moreover, I report coverage rates using bootstrapped confidence bands.

Table 2 reports the results. Three patterns stand out. First, for *Design 1* with the hump-shaped responses, the errors of the IR estimates are limited, and decreasing as the sample size increases. Second, the coverage rates are generally high, albeit lower than 95%. Third, the results deteriorate significantly moving to *Design 2*, especially for horizons 5 to 8. As a comparison, I presents an example of estimates in Figure 2. It is evident that the estimated IRs are more slow-moving than the truth. Hence FVAR may not be able to estimate responses when the true IRs experience sharp jumps.

6 Empirical application

In this section, I apply the functional FVAR approach to extend the results in Cloyne and Surico (2017) and Anderson et al. (2016) to study the impact of tax shocks on income inequality in the United Kingdom. I study tax policy for two reasons. First, it is generally considered to be an important driver of income inequality meanwhile the trade-off between economic growth and equality remains controversial (Coady and Gupta, 2012; Biswas et al., 2017). Second, with the narrative shocks constructed by Cloyne (2013), I can study the interaction of tax shocks and income distribution for four decades, which gives much greater variations in the data.

Specifically, I study the effects of tax policy changes in the United Kingdom from 1968Q1 to 2009Q4. The baseline model is the VARX model as in Cloyne (2013). As is shown in (36), $\mathbf{Y}_t$ is a vector of aggregate variables including the log of real per capita GDP, consumption and investment. $\mathbf{t}$ is the linear trend and the shock s_t is the tax policy changes based on long-run considerations. By definition, s_t should be exogenous to both shocks at business cycle frequency and equality concerns. Finally, $\mathbf{B(L)}$ and $\mathbf{D(L)}$ are lag polynomials with P and Q lags.

$$\mathbf{Y}_t = \mathbf{C}_0 + \mathbf{C}_1\mathbf{t} + \mathbf{B(L)}\mathbf{Y}_{t-1} + \mathbf{D(L)}s_t + \mathbf{e}_t \tag{36}$$

The specification is superior to SVAR-IV alternative (Mertens and Ravn, 2013; Stock and Watson, 2018; Plagborg-Møller and Wolf, 2021) as it allows for flexible lag structure, which is particularly useful when the sample size is limited. However, alternative specifications yield similar results.

Next model (36) is extended in two ways. First, I augment it with income density functions of the UK households following the FVAR procedure outlined in Section 2. Second, as a comparison I estimate the same model with commonly used inequality measures: 90-10 ratio, mean log deviation (MLD), Gini coefficient and Theil index. In all cases, I include four lags of the dependent variables and 12 lags of the tax shocks, but the results are robust to different lag lengths.

In what follows I first discuss the data source, income density estimation and the FPCA step in Section 6.1. Section 6.2 focuses on the SVAR analysis.

6.1 Functional analysis of inequality

The household income distributions are constructed using the Household Below Average Income (HBAI) dataset, which is based on the Family Expenditure Survey (FES) from 1968 to 1993 and the Family Resource Survey (FRS) from 1994 onwards.[7] The two surveys are the most comprehensive sources of household income in the United Kingdom, and have been extensively used to measure inequality (Goodman and Webb, 1994; Torry et al., 2019; Xu et al., 2019). On average the FES interviewed 2,270 households every quarter, which is expanded to 6,639 when FRS is adopted.[8] To estimate income distributions, I use household-level equivalized deflated weekly net income after deducting housing costs and adjusted for under-reporting of the rich households. For instance, an observation with £500 should be interpreted as a childless couple earning 500 pounds per week in 2018 prices. A brief overview of data preprocessing is provided in Appendix B.

With repeated cross-section survey data, I estimate the income density function from 1968Q1 to 2018Q4. The estimated densities are shown in Figure 3. As panel 3(a) shows, income distributions in the UK have experienced sizable fluctuations since 1968. Moreover, there is a continuing trend of rising inequality —with the current distributions having much fatter tails. Besides, the mean and median of the distributions also steadily shift upward, indicating that households are better off than they were fifty years ago. Overall, the estimated density functions are in line with previous findings (Cribb et al., 2016).

However, as Assumption 1 indicates, consistency for FPCA requires the functions to be

[7] The survey data are compiled and converted to harmonized income series by Institute for Fiscal Studies (2020) and Department for Work and Pensions (2020) for FES and FRS respectively.

[8] The data for 1991Q2 are dropped due to confusions caused by new council tax, and data for 1994Q1 are missing because of insufficient observations in the survey. For both periods, I interpolate income distributions from adjacent periods.

stationary. Therefore I detrend the density functions by removing a quadratic trend, which is presented in panels 3(b) and 3(c). As we can see, the fluctuations of densities remain sizable even after removing the trend. Furthermore, the fluctuations are moderated since 1990s, which coincides with the "Great Moderation" in the UK (Benati, 2008).

Next, I conduct functional principal component analysis to approximate the dynamics of income distributions. Figure 4 presents a set of criteria for selecting the number of functional principal components. First, the errors associated with FPCA decrease sharply with three FPCs, and the gain from over five FPCs is negligible. This is consistent with the results based on fraction of variations explained (FVE). The bottom panel shows that the first five FPCs explain 97.7% of the variations explained. Given that, the order of eigenfunctions is set to five. Finally, Figure 5 shows the density functions recovered by FPCA. As we can see, the FPCs successfully capture the fluctuations of the income distributions.

6.2 Tax cuts and income distributions

I now turn to study the interaction between tax shocks and income distributions. To start with, I report in Figure 6 the responses of aggregates in both the baseline and the FVAR model. Two observations stand out. First, the exogenous tax cuts are unambiguously expansionary: A one percentage point cut in taxes raises GDP by 2.6% in 10 quarters. Similarly, consumption and investment also increase persistently. Overall, the estimates are similar to Cloyne (2013). Second, including information about income distributions does not change the results. In fact, the responses of the aggregates are almost identical.

Next, I recover the responses of income densities in Figure 7. Three comments are in order. First, following a tax cut, there is a sizable decrease in the densities of households with weekly income between £100 and £300, compensated by an unambiguous increase in densities of the rich with weekly income above 300 pounds. The result suggests that the lower-middle-class households are better off after tax cuts.[9] One potential explanation is

[9] Note that median households income exceeded 300 pounds only after 1995Q1.

that tax cuts boost the economy and thus even poor households benefit through general equilibrium effects, e.g. stronger demand and higher wages (Cloyne, 2013).

Second, the increased density of the extremely poor (with weekly income below 50 pounds) does raise equality concerns for such policy. One potential channel is that tax cuts induce lower social securities which disproportionately hit the extremely poor. However, without detailed data on social security spending, it is difficult to test the channel.

Third, we may want to understand the contribution of tax shocks to the evolution of income distributions. To do so, I simulate the FVAR model by keeping only the tax shocks. Figure 8 contrasts the actual density dynamics with the simulated series. Perhaps surprisingly, the simulated densities follow the actual ones extremely even without other economic shocks. Moreover, the magnitude of the fluctuations is about half of the data. Overall, it suggests that tax shocks account for a non-negligible part of the income dynamics.

Taken together, the FVAR analysis demonstrates the distinct trade-off between economic booms and equality concerns of tax policy. Moreover, tax cuts are important in explaining the evolution of income distributions.

However, the above findings may be missed with conventional measures of inequality. To see that, I report in Figure 9 the responses of inequality measures using the same model. Clearly, all three inequality measures increase persistently following the expansionary tax shocks. We may therefore conclude that tax cuts are harmful for the poor households. However, as Figure 7(b) suggests, even though conventional inequality indicators increase due to the shrinking middle class, lower middle households in fact benefit from the cuts. Put it differently, standard inequality metrics provide little information on how different parts of the income distribution respond, and may result in misleading conclusions.

7 Conclusion

This paper introduces a method to estimate a VAR model with functional variables. Taking advantage of recent developments in functional principal component analysis (FPCA), the functional VAR model is able to reduce the dimension of functions while preserving valuable information. Moreover, the FVAR model can be easily estimated by three steps similar to the FAVAR model, and standard identification and estimation techniques can be applied.

In the empirical application, I apply the functional VAR approach to study the impact of tax shocks on income distributions in the United Kingdom. I find that exogenous tax cuts have heterogeneous effects on households depending on the income level. Even though the extremely poor households are hit adversely, a large fraction of lower middle class households benefit from the cuts. Moreover, tax policy changes account for a significant portion of the cyclical evolution in the income distributions. Importantly, the findings are ignored when density functions are replaced by conventional inequality measures.

The paper also raises several promising research avenues for the future. First, it is interesting to evaluate the relative forecasting performance of the proposed FVAR models with conventional small-scale VARs. Second, FVAR facilitates the study of "distributional shocks" in driving aggregate fluctuations. Both possibilities speak directly to the core implications of theoretical heterogeneous agent models and thus can provide valuable insights in macroeconomic modeling and policy making.

References

Anderson, E., Inoue, A., and Rossi, B. (2016). Heterogeneous consumers and fiscal policy shocks. *Journal of Money, Credit and Banking*, 48(8):1877–1888.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Benati, L. (2008). The "Great Moderation" in the united kingdom. *Journal of Money, Credit and Banking*, 40(1):121–147.

Benhabib, J., Bisin, A., and Zhu, S. (2011). The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica : journal of the Econometric Society*, 79(1):123–157.

Benko, M. (2007). Functional data analysis with applications in finance.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*. *Quarterly Journal of Economics*, 120(1):387–422.

Biswas, S., Chakraborty, I., and Hai, R. (2017). Income inequality, tax policy, and economic growth. *The Economic Journal*, 127(601):688–727.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Number 149 in Lecture Notes in Statistics. Springer, New York.

Cai, T. T. and Hall, P. (2006). Prediction in Functional Linear Regression. *The Annals of Statistics*, 34(5).

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.

Chang, M., Chen, X., and Schorfheide, F. (2022). Heterogeneity and Aggregate Fluctuations.

Clark, T. and Leicester, A. (2005). Inequality and two decades of British tax and benefit reforms. *Fiscal Studies*, 25(2):129–158.

Cloyne, J. (2013). Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom. *American Economic Review*, 103(4):1507–1528.

Cloyne, J. S. and Surico, P. (2017). Household Debt and the Dynamic Effects of Income Tax Changes. *The Review of Economic Studies*, 84(1):45–81.

Coady, D. and Gupta, M. S. (2012). *Income Inequality and Fiscal Policy*. International Monetary Fund.

Coibion, O., Gorodnichenko, Y., Kueng, L., and Silvia, J. (2017). Innocent Bystanders? Monetary policy and inequality. *Journal of Monetary Economics*, 88:70–89.

Colciago, A., Samarina, A., and Haan, J. (2019). Central Bank Policies and Income and Wealth Inequality: A Survey. *Journal of Economic Surveys*, 33(4):1199–1231.

Cribb, J., Belfield, C., Joyce, R., and Hood, A. (2016). Living standards, poverty and inequality in the UK: 2016. Technical report, IFS.

Department for Work and Pensions (2020). Households Below Average Income, 1994/95-2019/20. [data collection].

Erceg, C. J. and Lindé, J. (2013). Fiscal consolidation in a currency union: Spending cuts vs. tax hikes. *Journal of Economic Dynamics and Control*, 37(2):422–445.

Glomm, G., Jung, J., and Tran, C. (2018). Fiscal austerity measures: Spending cuts vs. Tax increases. *Macroeconomic Dynamics*, 22(2):501–540.

Goodman, A. and Webb, S. (1994). For richer, for poorer: The changing distribution of income in the UK, 1961-91. *Fiscal Studies*, 15(4):29–62.

Hall, P. and Horowitz, J. L. (2007). Methodology and Convergence Rates for Functional Linear Regression. *The Annals of Statistics*, 35(1).

Hall, P. and Hosseini-Nasab, M. (2006). On Properties of Functional Principal Components Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.

Hörmann, S. and Kidziński, Ł. (2014). A note on estimation in hilbertian linear models. *Scandinavian Journal of Statistics*, 42(1):43–62.

Hörmann, S. and Kokoszka, P. (2010). Weakly Dependent Functional Data. *The Annals of Statistics*, 38(3).

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer Verlag, New York, N.Y.

Huber, F., Marcellino, M., and Tornese, T. (2023). The Distributional Effects of Economic Uncertainty.

Inoue, A. and Rossi, B. (2021). A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy. *Quantitative Economics*, 12(4):1085–1138.

Institute for Fiscal Studies (2020). Households Below Average Income Dataset, 1961-1993. [data collection].

Jappelli, T. and Pistaferri, L. (2014). Fiscal policy and MPC heterogeneity. *American Economic Journal: Macroeconomics*, 6(4):107–136.

Keiller, A. N., Joyce, R., and Bourquin, P. (2020). Living standards, poverty and inequality in the UK: 2020. Technical report, The IFS.

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

Kokoszka, P. and Reimherr, M. (2013). Determining the Order of the Functional Autoregressive Model. *Journal of Time Series Analysis*, 34(1):116–129.

Li, Y., Wang, N., and Carroll, R. J. (2013). Selecting the Number of Principal Components in Functional Data. *Journal of the American Statistical Association*, 108(504):1284–1294.

Ljungqvist, A. and Smolyansky, a. M. (2016). To cut or not to cut? On the impact of corporate taxes on employment and income. *Finance and Economics Discussion Series*, 2016(006):1–60.

Mertens, K. and Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, 103(4):1212–1247.

Petersen, A. and Müller, H.-G. (2016). Functional Data Analysis for Density Functions by Transformation to a Hilbert Space. *The Annals of Statistics*, 44(1).

Plagborg-Møller, M. and Wolf, C. K. (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York, 2nd ed edition.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243.

Shin, H. (2009). Partial Functional Linear Regression. *Journal of Statistical Planning and Inference*, 139(10):3405–3418.

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.

Torry, M. et al. (2019). Static microsimulation research on Citizen's Basic Income for the UK: A personal summary and further reflections. Technical report, EUROMOD at the Institute for Social and Economic Research.

Xu, X., Waters, T., Cribb, J., and Bourquin, P. (2019). Living standards, poverty and inequality in the UK: 2019. Technical report, The IFS.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Appendix A Proofs

The proof consists of two parts. First, Section A.1 establishes the classical results of functional principal component analysis in the current setup. Second, Section A.2 provides the consistency of the FVAR estimator. Throughout, $\tilde{(\cdot)}$ indicates infeasible estimators with perfectly observed functions. For instance,

$$\tilde{\mu}(u) = \frac{1}{T}\sum_{t=1}^{T} f_t(u), \quad \tilde{\gamma}(u,v) = \frac{1}{T}\sum_{t=1}^{T} f_t(u)f_t(v) - \tilde{\mu}(u)\tilde{\mu}(v) \ . \tag{37}$$

Moreover, Section A.1 deals with general functions that satisfy Assumption 1-2 and I do not distinguish between f_t and W_t.

A.1 Proposition 1

Four lemmas are used to prove Proposition 1. Lemmas 1 and 2 are Lemmas 4.2 and 4.3 in Bosq (2000). The two lemmas bound the eigen-elements by the norm of covariance operators, which is further bounded by the Hilbert-Schmidt norm. To apply the lemmas, we use the fact that the covariance operator Γ is a symmetric, positive-definite Hilbert-Schmidt operator (Horváth and Kokoszka, 2012). Formally, we define Hilbert-Schmidt operators as follows.

Definition 1. An operator A is called Hilbert-Schmidt if it admits a decomposition with orthonormal bases $\{e_i\}$ and $\{v_i\}$ such that

$$A(f) = \sum_{i=1}^{\infty} a_i \langle f, e_i \rangle v_i$$

and $\sum_{i=1}^{\infty} a_i < \infty$ for all $f \in H$. Denote the space of Hilbert-Schmidt operators by $\mathcal{HS}$. We have norm $\|A\|_{\mathcal{HS}} = \sqrt{\sum_{j=1}^{\infty} a_i^2}$.

Lemma 3 is taken from Hörmann and Kokoszka (2010) where they extend standard consistency to the case with weakly dependent functions. With this lemma, we can focus on the difference between our estimator and the infeasible counterparts, e.g. $\hat{\gamma} - \tilde{\gamma}$. To bound the difference, I rely on Lemma 4 on the consistency of the density estimation, which is taken from Petersen and Müller (2016). Specifically, we have

Lemma 1 (Bounded Eigenvalues). *Let $A_0 \in \mathcal{A}$ be a compact linear operator on the Hilbert space with spectral decomposition*

$$A_0(f) = \sum_{i=1}^{\infty} a_{0,i} \langle f, e_{0,i} \rangle v_{0,i}. \tag{38}$$

Then

$$a_{0,i} = \min_{A \in \mathcal{A}_{i-1}} \|A - A_0\|_{op}, i \geq 1 \tag{39}$$

where $\mathcal{A}_{i-1} = \{A : A \in \mathcal{A}, dim A(f) \leq i-1\}$. Moreover, given another compact linear operator $A_1(f) = \sum_{i=1}^{\infty} a_{1,i} \langle f, e_{1,i} \rangle v_{1,i}$, we have

$$|a_{i,1} - a_{i,0}| \leq \|A_1 - A_0\|_{op}, i \geq 1. \tag{40}$$

Lemma 2 (Bounded Eigenfunctions). *For covariance operator Γ and eigen-elements such that $\Gamma(\xi_i) = \lambda_i \xi_i$, and the estimates $\tilde{\Gamma}(\tilde{\xi}_i) = \tilde{\lambda}_i \tilde{\xi}_i$, we have*

$$\|\xi_i - \tilde{\xi}_i\| \leq \alpha_i \|\Gamma - \tilde{\Gamma}\|_{op}. \tag{41}$$

Lemma 3 (Perfectly Observed Functions). *Under Assumption 1, we have*

$$\begin{aligned} \|\mu - \tilde{\mu}\| &= O_p\left(T^{-1/2}\right) \\ \|\gamma - \tilde{\gamma}\| &= O_p\left(T^{-1/2}\right) \end{aligned} \tag{42}$$

.

Lemma 4 (Consistent Density Estimation). *Under Assumption 2 with the modified kernel* *density estimator* (14) *satisfies* $\hat{f} > 0$, $\int_0^1 \hat{f}(x)dx = 1$, *we have*

$$\sup_{f\in\mathcal{F}} E\left(\|f-\hat{f}\|^2\right) = O(h^2 + (Nh)^{-1}) \tag{43}$$

whenever $h \to 0$ *and* $Nh \to \infty$ *as* $N \to \infty$. N *is the number of i.i.d. observations.*

Finally, for future usage I list the inequalities that are repeatedly used here:

$$Eab \leq \sqrt{Ea^2}\sqrt{Eb^2} \tag{44}$$

$$|ab-cd|^2 \leq 2a^2(b-d)^2 + 2d^2(a-c)^2 \tag{45}$$

$$(a+b)^{1/2} \leq a^{1/2} + b^{1/2} \tag{46}$$

$$\left[\int f(x)g(x)dx\right]^2 \leq \int f(x)^2dx \int g(x)^2dx \tag{47}$$

Proof. (Proposition 1)

For illustrative purpose, let us focus on the second line as the proofs of the first two lines are almost identical. Moreover, to simplify notation here I assume that $\tilde{\mu} = 0$, but the proofs does not change with non-zero means. In this case we have

$$\tilde{\gamma}(u,v) = \frac{1}{T}\sum_{t=1}^{T} f_t(u)f_t(v), \quad \hat{\gamma}(u,v) = \frac{1}{T}\sum_{t=1}^{T} \hat{f}_t(u)\hat{f}_t(v). \tag{48}$$

By triangle inequality,

$$\|\gamma-\hat{\gamma}\| = \|\gamma-\tilde{\gamma}+\tilde{\gamma}-\hat{\gamma}\| \leq \underbrace{\|\gamma-\tilde{\gamma}\|}_{O_p(T^{-1/2})} + \|\tilde{\gamma}-\hat{\gamma}\| \tag{49}$$

where the first term is $O_p(T^{-1/2})$ by Lemma 3. We only need to bound the second term on the RHS. Specifically,

$$TE\|\tilde{\gamma}-\hat{\gamma}\|^2$$

$$
\begin{aligned}
=&T \iint E\left[\frac{1}{T}\sum_{t=1}^{T}\underbrace{(f_t(u)f_t(v)-\hat{f}_t(u)\hat{f}_t(v))}_{\text{denoted by } X_t}\right]^2 dudv.\\
=&\frac{1}{T}\iint\Big(E[X_1^2]+E[X_1X_2]+\cdots+E[X_1X_T]+\dots\\
&\qquad E[X_TX_1]+E[X_TX_2]+\cdots+E[X_T^2]\Big)dudv\\
=&\frac{1}{T}\iint\Big(TE[X_1^2]+2(T-1)E[X_1X_2]+\cdots+2E[X_1X_T]\Big)dudv\\
\leq&\underbrace{\iint E[X_1^2]dudv}_{\text{term ①}}+2*1*\underbrace{\sum_{p=1}^{T-1}\iint E[X_1X_{1+p}]dudv}_{\text{term ②}}
\end{aligned} \tag{50}
$$

where the third equality comes from the assumption that X_t is m-dependent and stationary (e.g. $E[X_2X_3]=E[X_1X_2]$). Moreover, the expectation commutes with the integral operators.

Consider first term ①. We have

$$
\iint E[X_1^2]dudv = E\iint[f_1(u)f_1(v)-\hat{f}_1(u)\hat{f}_1(v)]^2dudv\ . \tag{51}
$$

Note that

$$
\begin{aligned}
&[f_1(u)f_1(v)-\hat{f}_1(u)\hat{f}_1(v)]^2\\
=&\Big(f_1(u)[f_1(v)-\hat{f}_1(v)]+[f_1(u)-\hat{f}_1(u)][\hat{f}_1(v)-f_1(v)]\\
&\qquad+[f_1(u)-\hat{f}_1(u)]f_1(v)\Big)^2\\
=&f_1(u)^2[f_1(v)-\hat{f}_1(v)]^2+[f_1(u)-\hat{f}_1(u)]^2[\hat{f}_1(v)-f_1(v)]^2\\
&+[f_1(u)-\hat{f}_1(u)]^2f_1(v)^2+2f_1(u)[f_1(v)-\hat{f}_1(v)][f_1(u)-\hat{f}_1(u)]f_1(v)\\
&-2f_1(u)[f_1(v)-\hat{f}_1(v)]^2[f_1(u)-\hat{f}_1(u)]\\
&+2[f_1(u)-\hat{f}_1(u)]^2[\hat{f}_1(v)-f_1(v)]f_1(v)\ .
\end{aligned} \tag{52}
$$

Integrating the first three terms,

$$\begin{aligned}
&\iint f_1(u)^2[f_1(v) - \hat{f}_1(v)]^2 dudv \\
=&\int f_1(u)^2 du \int [f_1(v) - \hat{f}_1(v)]^2 dv = \|f_1\|^2 \|f_1 - \hat{f}_1\|^2
\end{aligned}$$

and similarly

$$\iint [f_1(u) - \hat{f}_1(u)]^2 [\hat{f}_1(v) - f_1(v)]^2 dudv = \|f_1 - \hat{f}_1\|^4$$

$$\iint [f_1(u) - \hat{f}_1(u)]^2 f_1(v)^2 dudv = \|f_1\|^2 \|f_1 - \hat{f}_1\|^2$$

Next, the last two terms cancel out when taking integrals. Specifically,

$$\begin{aligned}
&-2\iint f_1(u)[f_1(v) - \hat{f}_1(v)]^2 [f_1(u) - \hat{f}_1(u)] dudv \\
=&2\int f_1(u)[\hat{f}_1(u) - f_1(u)] du \int [f_1(v) - \hat{f}_1(v)]^2 dv
\end{aligned}$$

whereas the last line is

$$\begin{aligned}
&2\iint [f_1(u) - \hat{f}_1(u)]^2 [\hat{f}_1(v) - f_1(v)] f_1(v) dudv \\
=&2\int f_1(u)[f_1(u) - \hat{f}_1(u)] du \int [f_1(v) - \hat{f}_1(v)]^2 dv
\end{aligned}$$

Hence the only remaining term is

$$\begin{aligned}
&2\iint f_1(u)[f_1(v) - \hat{f}_1(v)][f_1(u) - \hat{f}_1(u)] f_1(v) dudv \\
=&2\left(\int f_1(u)[f_1(u) - \hat{f}_1(u)] du\right)^2 \\
\leq&2\int f_1(u)^2 du \int [f_1(u) - \hat{f}_1(u)]^2 du = 2\|f_1\|^2 \|f_1 - \hat{f}_1\|^2
\end{aligned} \tag{53}$$

where we use the integral inequality (47). Combing all the terms in (51), we have

$$\begin{aligned}\iint E[X_1^2]dudv &\le E4\|f_1\|^2\|f_1-\hat{f}_1\|^2+\|f_1-\hat{f}_1\|^4\\ &= E\|f_1-\hat{f}_1\|^2(4\|f_1\|^2+\|f_1-\hat{f}_1\|^2). \end{aligned} \tag{54}$$

Next consider term ②. We first define an auxiliary

$$X_{1+p}^{(p)} = f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v) - \hat{f}_{1+p}^{(p)}(u)\hat{f}_{1+p}^{(p)}(v).$$

Then we have

$$\begin{aligned} E[X_1X_{1+p}] &= E[X_1(X_{1+p}-X_{1+p}^{(p)})] + \underbrace{E[X_1X_{1+p}^{(p)}]}_{=0} \\ &\le \sqrt{EX_1^2}\sqrt{E(X_{1+p}-X_{1+p}^{(p)})^2} \end{aligned} \tag{55}$$

Note that

$$\begin{aligned} &\left(X_{1+p}-X_{1+p}^{(p)}\right)^2 \\ =&\left(f_{1+p}(u)f_{1+p}(v) - f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v) + \hat{f}_{1+p}^{(p)}(u)\hat{f}_{1+p}^{(p)}(v) - \hat{f}_{1+p}(u)\hat{f}_{1+p}(v)\right)^2 \\ \le& 2\left(f_{1+p}(u)f_{1+p}(v) - f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v)\right)^2 \\ &\quad + 2\left(\hat{f}_{1+p}^{(p)}(u)\hat{f}_{1+p}^{(p)}(v) - \hat{f}_{1+p}(u)\hat{f}_{1+p}(v)\right)^2 \end{aligned} \tag{56}$$

So (55) can be expanded to two terms, using inequality (46):

$$\sqrt{EX_1^2}\sqrt{2E(f_{1+p}(u)f_{1+p}(v) - f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v))^2} \tag{57}$$

$$\sqrt{EX_1^2}\sqrt{2E(\hat{f}_{1+p}(u)\hat{f}_{1+p}(v) - \hat{f}_{1+p}^{(p)}(u)\hat{f}_{1+p}^{(p)}(v))^2} \tag{58}$$

The bounds of these terms are obtained in similar ways. Consider for instance (57).

Using inequality (45) and (46), we have

$$
\begin{aligned}
&\sqrt{EX_1^2}\sqrt{2E(f_{1+p}(u)f_{1+p}(v)-f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v))^2}\\
\le&\sqrt{EX_1^2}\sqrt{4Ef_{1+p}^2(u)(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2}\\
&\quad+\sqrt{EX_1^2}\sqrt{4E{f_{1+p}^{(p)}}^2(v)(f_{1+p}(u)-f_{1+p}^{(p)}(u))^2}.
\end{aligned}
$$

Therefore, summing over $(T-1)$ terms, we have

$$
\begin{aligned}
&\sum_{p=1}^{T-1}\iint\sqrt{EX_1^2}\sqrt{2E(f_{1+p}(u)f_{1+p}(v)-f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v))^2}dudv\\
\le&\sum_{p=1}^{\infty}\iint\sqrt{EX_1^2}\sqrt{2E(f_{1+p}(u)f_{1+p}(v)-f_{1+p}^{(p)}(u)f_{1+p}^{(p)}(v))^2}dudv\\
\le&2\sum_{p=1}^{\infty}\iint\sqrt{EX_1^2}\sqrt{Ef_{1+p}^2(u)(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2}dudv\\
&\quad+2\sum_{p=1}^{\infty}\iint\sqrt{EX_1^2}\sqrt{E{f_{1+p}^{(p)}}^2(v)(f_{1+p}(u)-f_{1+p}^{(p)}(u))^2}dudv
\end{aligned}
\tag{59}
$$

The two parts share the same upper bound. For instance, we have for the first part

$$
\begin{aligned}
&2\sum_{p=1}^{\infty}\iint\sqrt{EX_1^2}\sqrt{Ef_{1+p}^2(u)(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2}dudv\\
\le&2\sum_{p=1}^{\infty}\left[\iint EX_1^2dudv\right]^{1/2}\left[\iint Ef_{1+p}^2(u)(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2dudv\right]^{1/2}\\
=&2\sum_{p=1}^{\infty}\left[\iint EX_1^2dudv\right]^{1/2}\left[E\int f_{1+p}^2(u)du\int(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2dv\right]^{1/2}\\
\le&2\sum_{p=1}^{\infty}\left[\iint EX_1^2dudv\right]^{1/2}\left[E\left(\int f_{1+p}^2(u)du\right)^2\right]^{1/4}\left[E\left(\int(f_{1+p}(v)-f_{1+p}^{(p)}(v))^2dv\right)^2\right]^{1/4}\\
=&2\left[\iint EX_1^2dudv\right]^{1/2}\left[E\|f\|^4\right]^{1/4}\sum_{p=1}^{\infty}\left[E\|f_{1+p}-f_{1+p}^{(p)}\|\right]^{1/4}
\end{aligned}
\tag{60}
$$

where we have used (47) and (46). Importantly, the infinite sum on the RHS is finite by

Assumption 1. That is, the term with (57) is bounded by

$$4\Big[\iint EX_1^2 dudv\Big]^{1/2}\Big[E\|f\|^4\Big]^{1/4}\sum_{p=1}^{\infty}\Big[E\|f_{1+p}-f_{1+p}^{(p)}\|\Big]^{1/4}.$$

In exactly the same way we bound (58) by

$$4\Big[\iint EX_1^2 dudv\Big]^{1/2}\Big[E\|\hat{f}\|^4\Big]^{1/4}\sum_{p=1}^{\infty}\Big[E\|\hat{f}_{1+p}-\hat{f}_{1+p}^{(p)}\|\Big]^{1/4}.$$

To see the above bound is finite, note that by Lemma 4

$$\|\hat{f}\| = \|f\| + O(\sqrt{h^2+(Nh)^{-1}})\ . \tag{61}$$

When $h \to 0$ and $Nh \to \infty$ we have $\|\hat{f}\| = \|f\| + o(1)$. In this case the above bound is finite following Assumption 1.

Taken together, (55) is bounded by

$$\Big[\iint EX_1^2 dudv\Big]^{1/2} * M \tag{62}$$

where M is a finite constant given by

$$M = 4\bigg(\Big[E\|f\|^4\Big]^{1/4}\sum_{p=1}^{\infty}\Big[E\|f_{1+p}-f_{1+p}^{(p)}\|\Big]^{1/4} + \Big[E\|\hat{f}\|^4\Big]^{1/4}\sum_{p=1}^{\infty}\Big[E\|\hat{f}_{1+p}-\hat{f}_{1+p}^{(p)}\|\Big]^{1/4}\bigg)\ .$$

Combining the two terms ① and ②, we have

$$TE\|\tilde{\gamma}-\hat{\gamma}\|^2 \le \iint E[X_1^2]dudv + 2M\Big[\iint EX_1^2 dudv\Big]^{1/2}. \tag{63}$$

By (54) and Lemma 4, the RHS would be finite, with the order of magnitude depending on the number of observations $N(t)$ and the bandwidth h. Therefore,

$$\|\tilde{\gamma} - \hat{\gamma}\|^2 = O_p(T^{-1/2}) \tag{64}$$

and we prove the root-T consistency of the covariance function:

$$\|\gamma - \hat{\gamma}\|^2 = O_p(T^{-1/2}). \tag{65}$$

Finally, let us prove the last two lines. As Hörmann and Kokoszka (2010) show, both $\hat{\gamma}(\cdot,\cdot)$ and $\gamma(\cdot,\cdot)$ are Hilbert-Schmidt kernels, we have $\hat{\Gamma} - \Gamma$ is a Hilbert-Schmidt operator with kernel $\hat{\gamma}(u,v) - \gamma(u,v)$. Moreover, it is easy to verify that

$$\|\Gamma - \hat{\Gamma}\|^2_{\mathcal{HS}} = \|\gamma - \hat{\gamma}\|^2. \tag{66}$$

Now by Lemma 1, we have

$$E|\lambda_k - \hat{\lambda}_k|^2 \leq E\|\Gamma - \hat{\Gamma}\|^2_{op} \leq E\|\Gamma - \hat{\Gamma}\|^2_{\mathcal{HS}} \tag{67}$$

which implies the third line. Similarly, by Lemma 2, we have

$$E\|\xi_i - \hat{\xi}_i\|^2 = E\|\xi_i - \breve{\xi}_i\|^2 \leq \alpha_i E\|\Gamma - \tilde{\Gamma}\|^2_{op} \leq \alpha_i E\|\Gamma - \tilde{\Gamma}\|^2_{\mathcal{HS}} \tag{68}$$

which implies the last line.

■

A.2 Consistency

To facilitate the proof, it is useful to start with a simple scalar-on-function regression. Lemma 5 states that in our setup with weak dependency and estimated densities, the functional operator can still be consistently estimated. Then I break down Theorem 1 into four lemmas from Lemma 6 to 9.

Lemma 5. With a bit abuse of notation, consider the model

$$Z_t = A(W_t) + U_t. \tag{69}$$

where U_t is i.i.d. white noise. Under Assumption 1-3, we have

$$\|A - \hat{A}_{q_W}\|_{op} \xrightarrow{p} 0, \quad \text{with } \hat{A}_{q_W} = \sum_{j=1}^{q_W} \hat{\phi}_j \circ \frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)}{\hat{\rho}_j} \ . \tag{70}$$

Lemma 6.

$$\mathbf{vec}(\hat{\Psi}_{11}) - \mathbf{vec}(\Psi_{11}) \xrightarrow{p} 0 \tag{71}$$

Lemma 7.

$$\|\hat{\Psi}_{12} - \Psi_{12}\|_{op} \xrightarrow{p} 0 \tag{72}$$

Lemma 8.

$$\|\hat{\Psi}_{21} - \Psi_{21}\| \xrightarrow{p} 0 \tag{73}$$

Lemma 9.

$$\|\hat{\Psi}_{22} - \Psi_{22}\|_{op} \xrightarrow{p} 0 \tag{74}$$

Proof. **(Lemma 5)**

To start with, notice that

$$A = \sum_{j=1}^{\infty} \phi_j \circ \frac{\Gamma_{WZ}(\phi_j)}{\rho_j}. \tag{75}$$

We have

$$
\begin{aligned}
&\|\hat{A}_{q_W} - A\|_{op} \\
\leq & \left\| \sum_{j=1}^{q_W} \left(\hat{\phi}_j \circ \frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)}{\hat{\rho}_j} - \hat{\phi}_j \circ \frac{\Gamma_{WZ}(\hat{\phi}_j)}{\hat{\rho}_j} \right) \right\|_{op} \\
& + \left\| \sum_{j=1}^{q_W} \left(\hat{\phi}_j \circ \frac{\Gamma_{WZ}(\hat{\phi}_j)}{\hat{\rho}_j} - \hat{\phi}_j \circ \frac{\Gamma_{WZ}(\hat{\phi}_j)}{\rho_j} \right) \right\|_{op} \\
& + \left\| \sum_{j=1}^{q_W} \left(\hat{\phi}_j \circ \frac{\Gamma_{WZ}(\hat{\phi}_j)}{\rho_j} - \phi_j \circ \frac{\Gamma_{WZ}(\phi_j)}{\rho_j} \right) \right\|_{op} \\
& + \left\| \sum_{j=1}^{q_W} \phi_j \circ \frac{\Gamma_{WZ}(\phi_j)}{\rho_j} - \sum_{j=1}^{\infty} \phi_j \circ \frac{\Gamma_{WZ}(\phi_j)}{\rho_j} \right\|_{op} .
\end{aligned}
$$

Denote the four terms on the right hand side by M_1, M_2, M_3, M_4 respectively. We want to show that they are all $o_p(1)$.

First consider term M_1. We have

$$
\begin{aligned}
&\mathbb{P}\left(M_1^2 > \epsilon\right) \\
=&\mathbb{P}\left(\left\| \sum_{j=1}^{q_W} \langle \hat{\phi}_j, \hat{\phi}_j \rangle \frac{(\hat{\Gamma}_{WZ} - \Gamma_{WZ})(\hat{\phi}_j)}{\hat{\rho}_j} \right\|_{L2}^2 > \epsilon \right) \\
=&\mathbb{P}\left(\sum_{j=1}^{q_W} \left\| \frac{(\hat{\Gamma}_{WZ} - \Gamma_{WZ})(\hat{\phi}_j)}{\hat{\rho}_j} \right\|_{L2}^2 > \epsilon \right) \\
\leq&\mathbb{P}\left(\frac{1}{\hat{\rho}_{q_W}^2} \sum_{j=1}^{q_W} \left\| (\hat{\Gamma}_{WZ} - \Gamma_{WZ})(\hat{\phi}_j) \right\|_{L2}^2 > \epsilon \right) \\
\leq&\mathbb{P}\left(\frac{1}{\hat{\rho}_{q_W}^2} \sum_{j=1}^{\infty} \left\| (\hat{\Gamma}_{WZ} - \Gamma_{WZ})(\hat{\phi}_j) \right\|_{L2}^2 > \epsilon \right)
\end{aligned}
$$

$$\leq \mathbb{P}\left(m_T^2 \left\| \hat{\Gamma}_{WZ} - \Gamma_{WZ} \right\|_{op}^2 > \epsilon \right) \leq E\|\hat{\Gamma}_{WZ} - \Gamma_{WZ}\|_{op}^2 \frac{m_T^2}{\epsilon} .$$

Specifically, the first line comes from the definition of operator norm and the fact that $\{\hat{\phi}_j\}$ are othonormal. The second line uses Assumption 3(i) that eigenvalues are in decreasing order and the second last inequality uses Assumption 3(ii). Finally, the last inequality is obtained by the Markov inequality. Note that in the proof of Proposition 1, I show that $TE\|\hat{\Gamma}_f - \Gamma_f\|_{op}^2 \leq C$ where C is some constant. In the same way, we have

$$TE\|\hat{\Gamma}_{WZ} - \Gamma_{WZ}\|_{op}^2 \leq C \tag{76}$$

and thus

$$\mathbb{P}\left(M_1^2 > \epsilon\right) \leq \frac{C}{T}\frac{m_T^2}{\epsilon}. \tag{77}$$

Next consider M_2. We have

$$\begin{aligned}
&\mathbb{P}(M_2^2 > \epsilon) \\
\leq &\mathbb{P}\left(\sum_{j=1}^{q_W} \left\| \Gamma_{WZ}(\hat{\phi}_j) \left(\frac{1}{\hat{\rho}_j} - \frac{1}{\rho_j} \right) \right\|_{L2}^2 > \epsilon \right) \\
\leq &\mathbb{P}\left(\max_{1\leq j\leq q_W} \left(\frac{\hat{\rho}_j - \rho_j}{\hat{\rho}_j \rho_j} \right)^2 \sum_{j=1}^{\infty} \|\Gamma_{WZ}(\hat{\phi}_j)\|^2 > \epsilon \right) \\
\leq &\mathbb{P}\left(\frac{1}{\hat{\rho}_{q_W}^2} \max_{1\leq j\leq q_W} \left(\frac{\hat{\rho}_j - \rho_j}{\rho_j} \right)^2 \|\Gamma_{WZ}\|_{op}^2 > \epsilon \right) \\
\leq &\mathbb{P}\left(\max_{1\leq j\leq q_W} \left| \frac{\hat{\rho}_j - \rho_j}{\rho_j} \right| > \sqrt{\frac{\epsilon}{m_T^2 \|\Gamma_{WZ}\|_{op}^2}} \right) \\
\leq &\mathbb{P}\left(\max_{1\leq j\leq q_W} |\hat{\rho}_j - \rho_j| > \frac{1}{2m_T} \sqrt{\frac{\epsilon}{m_T^2 \|\Gamma_{WZ}\|_{op}^2}} \right) \\
&+ \mathbb{P}\left(\left\{ \frac{1}{\rho_{q_W}} \max_{1\leq j\leq q_W} |\hat{\rho}_j - \rho_j| > \sqrt{\frac{\epsilon}{m_T^2 \|\Gamma_{WZ}\|_{op}^2}} \right\} \right. \\
&\qquad\qquad \left. \bigcap \left\{ \max_{1\leq j\leq q_W} |\hat{\rho}_j - \rho_j| \leq \frac{1}{2m_T} \sqrt{\frac{\epsilon}{m_T^2 \|\Gamma_{WZ}\|_{op}^2}} \right\} \right)
\end{aligned}$$

The derivation uses the same trick as before. Consider the first term in the last inequality

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le q_W}|\hat{\rho}_j-\rho_j|>\frac{1}{2m_T}\sqrt{\frac{\epsilon}{m_T^2\|\Gamma_{WZ}\|_{op}^2}}\right)\\
\le&\mathbb{P}\left(\|\hat{\Gamma}_f-\Gamma_f\|_{op}>\frac{1}{2m_T}\sqrt{\frac{\epsilon}{m_T^2\|\Gamma_{WZ}\|_{op}^2}}\right)\\
\le&E\|\hat{\Gamma}_f-\Gamma_f\|_{op}^2\frac{4m_T^4\|\Gamma_{WZ}\|_{op}^2}{\epsilon}\\
\le&\frac{C}{T}\frac{4m_T^4\|\Gamma_{WZ}\|_{op}^2}{\epsilon}
\end{aligned}$$

where again we use Markov inequality and Proposition 1. For the second term we note that it is upper bounded by

$$\mathbb{P}\left(\frac{1}{2m_T}>\rho_{q_W}\bigcap|\hat{\rho}_{q_W}-\rho_{q_W}|\le\frac{1}{2m_T}\sqrt{\frac{\epsilon}{m_T^2\|\Gamma_{WZ}\|_{op}^2}}\right)\xrightarrow{p}0\ . \tag{78}$$

To see that, notice that $\sqrt{\frac{\epsilon}{m_T^2\|\Gamma_{WZ}\|_{op}^2}}$ goes to zero because ϵ is an arbitrarily small number while $m_T\to\infty$ as T goes to infinity. Moreover, by Assumption 3(ii) we have $\hat{\rho}_{q_W}>\frac{1}{m_T}$ and we require $|\hat{\rho}_{q_W}-\rho_{q_W}|$ shrinks at a rate faster than $\frac{1}{2m_T}$. In this case $\rho_{q_W}>\frac{1}{2m_T}$ and thus the above probability converges to zero. Taken together, we have

$$\mathbb{P}(M_2^2>\epsilon)\le\frac{C}{T}\frac{4m_T^4\|\Gamma_{WZ}\|_{\mathcal{HS}}^2}{\epsilon}. \tag{79}$$

As for M_3, note that

$$\hat{\phi}_j\circ\Gamma_{WZ}(\hat{\phi}_j)-\phi_j\circ\Gamma_{WZ}(\phi_j)=\hat{\phi}_j\circ\Gamma_{WZ}(\hat{\phi}_j-\phi_j)+(\hat{\phi}_j-\phi_j)\circ\Gamma_{WZ}(\phi_j). \tag{80}$$

Therefore,

$$\mathbb{P}(M_3>\epsilon)$$

$$
\begin{aligned}
=&\mathbb{P}\left(\left\|\sum_{j=1}^{q_W}\frac{1}{\rho_j}\left(\hat{\phi}_j\circ\Gamma_{WZ}(\hat{\phi}_j-\phi_j)+(\hat{\phi}_j-\phi_j)\circ\Gamma_{WZ}(\phi_j)\right)\right\|_{op}>\epsilon\right)\\
\leq&\mathbb{P}\left(\sum_{j=1}^{q_W}\frac{1}{\rho_j}\left(\left\|\hat{\phi}_j\circ\Gamma_{WZ}(\hat{\phi}_j-\phi_j)\right\|_{op}+\left\|(\hat{\phi}_j-\phi_j)\circ\Gamma_{WZ}(\phi_j)\right\|_{op}\right)>\epsilon\right)\\
\leq&\mathbb{P}\left(\sum_{j=1}^{q_W}\frac{2}{\rho_j}\|\hat{\phi}_j-\phi_j\|_{L2}\|\Gamma_{WZ}\|_{op}>\epsilon\right)\\
\leq&\mathbb{P}\left(\frac{1}{\rho_{q_W}}\sum_{j=1}^{q_W}\|\hat{\phi}_j-\phi_j\|_{L2}>\frac{\epsilon}{2\|\Gamma_{WZ}\|_{op}}\right)\\
\leq&\mathbb{P}\left(\sum_{j=1}^{q_W}\|\hat{\phi}_j-\phi_j\|_{L2}>\frac{\epsilon}{4m_T\|\Gamma_{WZ}\|_{op}}\right)+\mathbb{P}\left(\frac{1}{\rho_{q_W}}>2m_T\right). \qquad (81)
\end{aligned}
$$

The first two lines above are straightforward. The third line is obtained by the definition of operator norm. For example, we have

$$
\begin{aligned}
\left\|\hat{\phi}_j\circ\Gamma_{WZ}(\hat{\phi}_j-\phi_j)\right\|_{op}=&\sup\left\{\left\|\langle\hat{\phi}_j,g\rangle\Gamma_{WZ}(\hat{\phi}_j-\phi_j)\right\|_{L2},\forall\|g\|\leq 1\right\}\\
=&\|\Gamma_{WZ}(\hat{\phi}_j-\phi_j)\|_{L2}\leq\|\Gamma_{WZ}\|_{op}\|\hat{\phi}_j-\phi_j\|_{L2}\ . \qquad (82)
\end{aligned}
$$

The last two inequalities use the same trick as before. Next we have

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{q_W}\|\hat{\phi}_j-\phi_j\|_{L2}>\frac{\epsilon}{4m_T\|\Gamma_{WZ}\|_{op}}\right)\\
\leq&\mathbb{P}\left(m_T\max_{1\leq j\leq R_2}\|\hat{\phi}_j-\phi_j\|_{L2}>\frac{\epsilon}{4m_T\|\Gamma_{WZ}\|_{op}}\right)\\
\leq&\mathbb{P}\left(m_T\max_{1\leq j\leq R_2}\frac{2\sqrt{2}}{\hat{\alpha}_j}\|\hat{\Gamma}_f-\Gamma_f\|_{op}>\frac{\epsilon}{4m_T^2\|\Gamma_{WZ}\|_{op}}\right)\\
\leq&\mathbb{P}\left(\|\hat{\Gamma}_f-\Gamma_f\|_{op}>\frac{\epsilon}{8\sqrt{2}m_T^3\|\Gamma_{WZ}\|_{op}}\right)\\
\leq&E\|\hat{\Gamma}_f-\Gamma_f\|_{op}^2128\|\Gamma_{WZ}\|_{op}^2\frac{m_T^6}{\epsilon^2}\leq\frac{C}{T}\frac{128\|\Gamma_{WZ}\|_{op}^2m_T^6}{\epsilon^2} \qquad (83)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{\rho_{qW}} > 2m_T\right) \\
\leq &\underbrace{\mathbb{P}\left(\frac{1}{\rho_{qW}} > 2m_T \bigcap |\hat{\rho}_{qW} - \rho_{qW}| \leq \frac{1}{2m_T}\right)}_{=0} + \mathbb{P}\left(|\hat{\rho}_{qW} - \rho_{qW}| > \frac{1}{2m_T}\right) \\
\leq &\mathbb{P}\left(\|\hat{\Gamma}_f - \Gamma_f\|_{op} > \frac{1}{2m_T}\right) \leq E\|\hat{\Gamma}_f - \Gamma_f\|_{op}^2 4m_T^2 \leq \frac{C}{T}4m_T^2. \qquad (84)
\end{aligned}
$$

Combing the two terms, we have

$$
\mathbb{P}(M_3 > \epsilon) \leq \frac{C}{T}\left(\frac{128\|\Gamma_{WZ}\|_{op}^2 m_T^6}{\epsilon^2} + 4m_T^2\right). \qquad (85)
$$

Finally let us bound M_4. Since the operator is bounded. Therefore, for an arbitrarily small positive number ϵ, we can always find a q_ϵ that is large enough such that

$$
\|A - A_{q_\epsilon}\|_{op}^2 = \sum_{j>q_\epsilon} \|A(\nu_j)\|_{L2}^2 < \epsilon. \qquad (86)
$$

Given that, we have

$$
\mathbb{P}\left(\|A - A_{qW}\|_{op}^2 > \epsilon\right) \leq \mathbb{P}\left(q_W \leq q_\epsilon\right). \qquad (87)
$$

Now combine the bounds (77),(79),(85) and (87), we have

$$
\begin{aligned}
&\mathbb{P}\left(\|\hat{A}_{qW} - A\|_{op} > \epsilon\right) \\
\leq &\frac{C}{T}\left(\frac{m_T^2}{\epsilon} + \frac{4m_T^4\|\Gamma_{WZ}\|_{op}^2}{\epsilon} + \frac{128\|\Gamma_{WZ}\|_{op}^2 m_T^6}{\epsilon^2} + 4m_T^2\right) + \mathbb{P}\left(q_W \leq q_\epsilon\right)
\end{aligned}
$$

which is $o_p(1)$ if $m_T^6 = o_p(T)$ and $q_W \to \infty$. ■

Proof. (**Lemma 6**)

Denote $\Psi_{12}(W) = [\Psi_{12}(W_1), \ldots, \Psi_{12}(W_T)]$ a $K \times T$ matrix. Then $Y = \Psi_{12}Z + \Psi_{12}(W) + e_y$.

We have

$$
\begin{aligned}
\hat{\Psi}_{11} &= [\Psi_{12}Z + \Psi_{12}(W) + e_y]\hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1} \\
&= \Psi_{12} + \left(\frac{1}{T}\sum_{t=1}^{T}\Psi_{12}(W_t)\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right]\right)\left(\frac{Z\hat{M}_\pi Z'}{T}\right)^{-1} \\
&\quad + \left(\frac{1}{T}\sum_{t=1}^{T}e_{yt}\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right]\right)\left(\frac{Z\hat{M}_\pi Z'}{T}\right)^{-1} \qquad (88)
\end{aligned}
$$

where $\hat{M}_\pi = I - \hat{\pi}'(\hat{\pi}\hat{\pi}')^{-1}\hat{\pi}$ is the residual maker.

Consistency requires that

$$
\frac{Z\hat{M}_\pi Z'}{T} = \left(\hat{\Gamma}_Z - \sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right) \xrightarrow{p} \Omega_{Z\pi Z} \qquad (89)
$$

where $\Omega_{Z\pi Z}$ is

$$
\Omega_{Z\pi Z} = \Gamma_Z - \sum_{j=1}^{\infty}\frac{\Gamma_{WZ}(\phi_j)\Gamma_{WZ}(\phi_j)'}{\rho_j} \qquad (90)
$$

which is positive definite, together with

$$
\frac{1}{T}\sum_{t=1}^{T}\Psi_{12}(W_t)\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right] = o_p(1) \qquad (91)
$$

and

$$
\frac{1}{T}\sum_{t=1}^{T}e_{yt}\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right] = o_p(1). \qquad (92)
$$

The proof hence proceeds in three steps.

Step 1. Notice that

$$
\begin{aligned}
&\left\|\hat{\Gamma}_Z - \sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j} - \left(\Gamma_Z - \sum_{j=1}^{\infty}\frac{\Gamma_{WZ}(\phi_j)\Gamma_{WZ}(\phi_j)'}{\rho_j}\right)\right\| \\
\leq &\|\hat{\Gamma}_Z - \Gamma_Z\| + \left\|\sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j} - \sum_{j=1}^{\infty}\frac{\Gamma_{WZ}(\phi_j)\Gamma_{WZ}(\phi_j)'}{\rho_j}\right\|
\end{aligned}
$$

$$=\|\hat{\Gamma}_Z - \Gamma_Z\| + \left\| A_{q_W}(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t') - A(EW_t Z_t')\right\| \tag{93}$$

where the second line comes from the triangle inequality and the third line uses the definition of A_{q_W} defined in Lemma 5. To see this, note that

$$A_{q_W}(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t') = \sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\langle\hat{\phi}_j, \frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t'\rangle}{\hat{\rho}_j} = \sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\frac{1}{T}\sum_{t=1}^{T}\langle\hat{\phi}_j, \hat{W}_t\rangle Z_t'}{\hat{\rho}_j} .$$

We study the two terms above separately. The first term is o_p since

$$\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t' \xrightarrow{p} EZ_t Z_t' \tag{94}$$

by law of large numbers. By add and subtract, the second term becomes

$$\begin{aligned}
&\Bigg\|(A_{q_W} - A)(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t' - EW_t Z_t') + \\
&\qquad (A_{q_W} - A)(EW_t Z_t') + A(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t' - EW_t Z_t')\Bigg\| \\
\leq & \left\|(A_{q_W} - A)(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t' - EW_t Z_t')\right\| + \left\|(A_{q_W} - A)(EW_t Z_t')\right\| \qquad (95)\\
&+ \left\|A(\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t' - EW_t Z_t')\right\| \\
\leq & (\|A_{q_W} - A\|_{op} + \|A\|_{op})\|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_t' - EW_t Z_t'\| + \|A_{q_W} - A\|_{op}\|EW_t Z_t'\| .
\end{aligned}$$

By Lemma 5, $\|A_{q_W} - A\|_{op} \xrightarrow{p} 0$. Moreover, following the proof in Proposition 1, we have for $l = 1, \ldots, Kp$

$$TE\|\frac{1}{T}\sum_{t=1}^{T}\hat{W}_t Z_{t,l} - EW_t Z_{t,l}\|_{L2}^2 \leq C \tag{96}$$

where C is some finite constant. Hence $\|\sum_{t=1}^{T} W_t Z_t' - EW_t Z_t'\| \xrightarrow{p} 0$. Taken together, we have

$$\hat{\Gamma}_Z - \sum_{j=1}^{q_W} \frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j} \xrightarrow{p} \Omega_{Z\pi Z}. \tag{97}$$

Step 2. Note that Ψ_{12} is a bounded linear operator, therefore,

$$\Psi_{12}(W)\hat{M}_\pi Z' = \Psi_{12}(\hat{W})\hat{M}_\pi Z' + \Psi_{12}(W - \hat{W})\hat{M}_\pi Z' \ . \tag{98}$$

Consider the first term, we have

$$\Psi_{12}(\hat{W}_t) = \Psi_{12}(\sum_{l=1}^{\infty} \hat{\pi}_{tl}\hat{\phi}_l) = \sum_{l=1}^{q_W} \Psi_{12}(\hat{\phi}_l)\hat{\pi}_{tl} + \sum_{l=q_W+1}^{\infty} \Psi_{12}(\hat{\phi}_l)\hat{\pi}_{tl} \ . \tag{99}$$

Multiply it by $\hat{M}_\pi Z'/T$, it becomes

$$\frac{1}{T}\Psi_{12}(\hat{W})\hat{M}_\pi Z' = \sum_{l=q_W+1}^{\infty} \frac{1}{T}\sum_{t=1}^{T} \Psi_{12}(\hat{\phi})\hat{\pi}_{tl} \left[Z_t' - \sum_{j=1}^{q_W} \frac{\langle \hat{W}_t, \hat{\phi}_j \rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j} \right] \tag{100}$$

where the first q_W terms disappear because $\hat{\pi}\hat{M}_\pi = 0$ by construction. Note also that FPC scores are mutually independent, thus

$$\frac{1}{T}\sum_{l=q_W+1}^{\infty} \sum_{j=1}^{q_W} \hat{\rho}_j^{-1}\hat{\pi}_{tl}\hat{\pi}_{tj}\hat{\Gamma}_{WZ}(\hat{\phi}_j)' = 0 \ . \tag{101}$$

The remaining terms

$$\sum_{l=q_W+1}^{\infty} \Psi_{12}(\hat{\phi})\frac{1}{T}\sum_{t=1}^{T} \hat{\pi}_{tl}Z_t' \xrightarrow{p} \sum_{l=q_W+1}^{\infty} \Psi_{12}(\hat{\phi})E\pi_{tl}Z_t' \leq \sqrt{E\pi_{tl}^2}\sqrt{EZ_t'Z_t} \ . \tag{102}$$

Since Y_t and thus Z_t is stationary, $EZ_t'Z_t$ is finite. In contrast, by Assumption 3 we have $E\pi_{tl}^2 = \lambda_l$ which converges to zero as $q_W \to \infty$.

Turning to $\Psi_{12}(W-\hat{W})\hat{M}_\pi Z'$, the logic follows similarly. Note that

$$\frac{1}{T}\Psi_{12}(W-\hat{W})\hat{M}_\pi Z' = \frac{1}{T}\sum_{t=1}^{T}\Psi_{12}(W_t-\hat{W}_t)\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right] . \tag{103}$$

Consider for example the product with Z_t', we have

$$\frac{1}{T}\sum_{t=1}^{T}\Psi_{12}(W_t-\hat{W}_t)Z_t' = \Psi_{12}\left(\frac{1}{T}\sum_{t=1}^{T}(W_t-\hat{W}_t)Z_t'\right) \xrightarrow{p} \Psi_{12}E(W_t-\hat{W}_t)Z_t' . \tag{104}$$

By Cauchy-SchwarZ inequality and Lemma 4, the right hand side is $o_p(1)$.

Step 3. Given what we have shown in Step 2, the last term is trivial to bound. Since

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}e_{yt}\left[Z_t' - \sum_{j=1}^{q_W}\frac{\langle \hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right] \\
=&\frac{1}{T}\sum_{t=1}^{T}e_{yt}Z_t' - \sum_{j=1}^{q_W}\frac{\langle \frac{1}{T}\sum_{t=1}^{T}e_{yt}\hat{W}_t, \hat{\phi}_j\rangle \hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}
\end{aligned} \tag{105}$$

and by Assumption 4(ii) e_{yt} is independent of both Z_t' and W_t. Hence the sample mean converges to zero.

Combining the results from the above steps, consistency follows by Slutsky theorem. ■

Proof. **(Lemma 7)**

Since the population model of the scalar block can be written as

$$Y = \Psi_{11}Z + \Psi_{12}(\hat{W}) + e_y + \Psi_{12}(W) - \Psi_{12}(\hat{W}) , \tag{106}$$

the estimator of Ψ_{12} is

$$\begin{aligned}
\hat{\Psi}_{12} =&[\Psi_{11}Z + \Psi_{12}(\hat{W}) + e_y + \Psi_{12}(W) - \Psi_{12}(\hat{W})]M_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1} \\
=&\Psi_{12}(\hat{W})M_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1} + e_yM_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1} \\
&+ [\Psi_{12}(W) - \Psi_{12}(\hat{W})]M_Z\hat{\pi}'(\hat{\pi}M_Z\hat{\pi}')^{-1}
\end{aligned}$$

Notice that

$$\begin{aligned}\Psi_{12}(\hat{W}) =& \left[\sum_{l=1}^{q_W} \hat{\pi}_{1,l}\Psi_{12}(\hat{\phi}_l), \ldots, \sum_{l=1}^{q_W} \hat{\pi}_{T,l}\Psi_{12}(\hat{\phi}_l)\right] \\ &+ \left[\sum_{l=q_W+1}^{\infty} \hat{\pi}_{1,l}\Psi_{12}(\hat{\phi}_l), \ldots, \sum_{l=q_W+1}^{\infty} \hat{\pi}_{T,l}\Psi_{12}(\hat{\phi}_l)\right] \\ =& \Psi_{12}(\hat{\phi})\hat{\pi} + \left[\sum_{l=q_W+1}^{\infty} \hat{\pi}_{1,l}\Psi_{12}(\hat{\phi}_l), \ldots, \sum_{l=q_W+1}^{\infty} \hat{\pi}_{T,l}\Psi_{12}(\hat{\phi}_l)\right] \end{aligned} \tag{107}$$

where $\Psi_{12}(\hat{\phi}) = [\Psi_{12}(\hat{\phi}_1), \ldots, \Psi_{12}(\hat{\phi}_{q_W})]$ is a $K \times q_W$ matrix. Denote $\hat{\Omega}_{\pi Z\pi} = \frac{1}{T}\hat{\pi}M_Z\hat{\pi}'$, and its *j-th* column as $\hat{\Omega}_{\pi Z\pi}[\cdot, j]$. Then $\sum_{j=1}^{q_W} \hat{\Psi}_{12}[\cdot, j]\hat{\phi}_j - \psi_{12}$ contains four terms:

$$\left(\sum_{j=1}^{q_W} \langle\hat{\phi}_j, \psi_{12}\rangle\hat{\phi}_j - \psi_{12}\right) = \sum_{j=q_W+1}^{\infty} \langle\hat{\phi}_j, \psi_{12}\rangle\hat{\phi}_j \tag{108}$$

$$\sum_{j=1}^{q_W} \frac{\sum_{t=1}^{T}\sum_{l=q_W+1}^{\infty} \hat{\pi}_{t,l}\Psi_{12}(\hat{\phi}_l)}{T} \left[\hat{\pi}_t' - Z_t'\hat{\Gamma}_Z\hat{\Gamma}_{WZ}(\hat{\phi})\right] \hat{\Omega}_{\pi Z\pi}^{-1}[\cdot, j]\hat{\phi}_j \tag{109}$$

$$\sum_{j=1}^{q_W} \frac{\sum_{t=1}^{T} \Psi_{12}(W_t - \hat{W}_t)}{T} \left[\hat{\pi}_t' - Z_t'\hat{\Gamma}_Z\hat{\Gamma}_{WZ}(\hat{\phi})\right] \hat{\Omega}_{\pi Z\pi}^{-1}[\cdot, j]\hat{\phi}_j \tag{110}$$

$$\sum_{j=1}^{q_W} \frac{\sum_{t=1}^{T} e_{yt}}{T} \left[\hat{\pi}_t' - Z_t'\hat{\Gamma}_Z\hat{\Gamma}_{WZ}(\hat{\phi})\right] \hat{\Omega}_{\pi Z\pi}^{-1}[\cdot, j]\hat{\phi}_j \tag{111}$$

The proof follows exactly the same logic as in Lemma 6, with the only difference that now the difference converges to zero in L^2 sense. Consider, for example the first term. Since ψ_{12} is in $L2$, we have

$$\|\psi_{12}\|^2 = \sum_{j=1}^{\infty} \langle\psi_{12}, \hat{\phi}_j\rangle^2 < \infty \tag{112}$$

by the Parseval identity. Given that it is bounded, we can always find a q_W that is large enough so that

$$\sum_{j=q_W+1}^{\infty} \langle\hat{\phi}_j, \psi_{12}\rangle \xrightarrow{p} 0. \tag{113}$$

In this case,

$$\| \sum_{j=q_W+1}^{\infty} \langle \hat{\phi}_j, \psi_{12} \rangle \hat{\phi}_j \| \xrightarrow{p} 0 \ . \tag{114}$$

The remaining terms are bounded in the same way. ■

Proof. **(Lemma 8)**

As (25) indicates, the functional coefficients are estimated by

$$\underbrace{\hat{\Psi}_{21}(u)}_{1\times Kp} = \sum_{j=1}^{q_f} \hat{\Psi}_{21}[j,\cdot]\hat{\xi}_j \tag{115}$$

where $\hat{\Psi}_{21}[j,\cdot]$ is the j-th row of

$$\hat{\Psi}_{21} = \hat{\eta}\hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1} \ . \tag{116}$$

We break down $\hat{\eta}$ into four terms. Note that η is a $q_f \times T$ matrix, with the (t,r) element

$$\begin{aligned} \hat{\eta}_{tr} &= \langle \hat{f}_t, \hat{\xi}_r \rangle = \langle f_t, \hat{\xi}_r \rangle + \langle \hat{f}_t - f_t, \hat{\xi}_r \rangle \\ &= \langle \Psi_{21} Z_t + \Psi_{22}(W_t) + e_{ft}, \hat{\xi}_r \rangle + \langle \hat{f}_t - f_t, \hat{\xi}_r \rangle \\ &= \langle \Psi_{21}, \hat{\xi}_r \rangle Z_t + \langle \Psi_{22}(W_t), \hat{\xi}_r \rangle + \langle e_{ft}, \hat{\xi}_r \rangle + \langle \hat{f}_t - f_t, \hat{\xi}_r \rangle \ . \end{aligned} \tag{117}$$

Hence $\hat{\eta}$ can be written as

$$\hat{\eta} = \langle \underbrace{\hat{\xi}}_{q_f\times 1}, \underbrace{\Psi_{21}}_{1\times Kp} \rangle \underbrace{Z}_{Kp\times T} + \langle \hat{\xi}, \underbrace{\Psi_{22}(W)}_{1\times T} \rangle + \langle \hat{\xi}, \underbrace{e_f}_{1\times T} \rangle + \langle \hat{\xi}, \underbrace{\hat{f} - f}_{1\times T} \rangle \ . \tag{118}$$

With the above decomposition, we can write $\hat{\Psi}_{21}$ as

$$\begin{aligned} \hat{\Psi}_{21} =& \langle \hat{\xi}, \Psi_{21} \rangle + \langle \hat{\xi}, \Psi_{22}(W) \rangle \hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1} \\ &+ \langle \hat{\xi}, e_f \rangle \hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1} + \langle \hat{\xi}, \hat{f} - f \rangle \hat{M}_\pi Z'(Z\hat{M}_\pi Z')^{-1} \ . \end{aligned} \tag{119}$$

Subtract Ψ_{21} from $\sum_{j=1}^{q_f} \hat{\Psi}_{21}[j,\cdot]\hat{\xi}_j(u)$, we have again four terms:

$$\left(\sum_{j=1}^{q_f}\langle\hat{\xi}_j,\Psi_{21}\rangle\hat{\xi}_j - \Psi_{21}\right) \tag{120}$$

$$\left(\frac{1}{T}\sum_{j=1}^{q_f}\langle\hat{\xi}_j,\Psi_{22}(W)\rangle\hat{\xi}_j\hat{M}_\pi Z'\right)\left(\frac{Z\hat{M}_\pi Z'}{T}\right)^{-1} \tag{121}$$

$$\left(\frac{1}{T}\sum_{j=1}^{q_f}\langle\hat{\xi}_j,\hat{f}-f\rangle\hat{\xi}_j\hat{M}_\pi Z'\right)\left(\frac{Z\hat{M}_\pi Z'}{T}\right)^{-1} \tag{122}$$

$$\left(\frac{1}{T}\sum_{j=1}^{q_f}\langle\hat{\xi}_j,e_f\rangle\hat{\xi}_j\hat{M}_\pi Z'\right)\left(\frac{Z\hat{M}_\pi Z'}{T}\right)^{-1} \tag{123}$$

Term 1.

Rewrite the first term by projecting Ψ_{21} on $\hat{\xi}_j$,

$$\left\|\sum_{j=1}^{q_f}\langle\hat{\xi}_j,\Psi_{21}\rangle\hat{\xi}_j - \sum_{j=1}^{\infty}\langle\hat{\xi}_j,\Psi_{21}\rangle\hat{\xi}_j\right\| = \left\|\sum_{j=q_f+1}^{\infty}\langle\hat{\xi}_j,\Psi_{21}\rangle\hat{\xi}_j\right\| \tag{124}$$

which is $o_p(1)$ following the logic of Lemma 7.

Term 2.

As for the second term, first notice that by Lemma 6,

$$\frac{Z\hat{M}_\pi Z'}{T} = \left(\hat{\Gamma}_Z - \sum_{j=1}^{q_W}\frac{\hat{\Gamma}_{WZ}(\hat{\phi}_j)\hat{\Gamma}_{WZ}(\hat{\phi}_j)'}{\hat{\rho}_j}\right) \xrightarrow{p} \Omega_{Z\pi Z} \tag{125}$$

which is positive definite. Hence we only discuss the converges of the left part. Consider the *j-th* item. It is equivalent to

$$\frac{1}{T}\sum_{t=1}^{T}\langle\hat{\xi}_j,\Psi_{22}(W_t)\rangle Z_t' - \sum_{l=1}^{q_W}\hat{\rho}_l^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\langle\hat{\xi}_j,\Psi_{22}(W_t)\rangle\hat{\pi}_{tl}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\pi}_{tl}Z_t'\right) \ . \tag{126}$$

Project W_t into the space of $\hat{\phi}_j$, the second part becomes

$$\begin{aligned}
&\sum_{l=1}^{q_W} \hat{\rho}_l^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \langle \hat{\xi}_j, \Psi_{22}(\sum_{r=1}^{\infty} \hat{\pi}_{tr} \hat{\phi}_r) \rangle \hat{\pi}_{tl} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \\
=& \sum_{l=1}^{q_W} \sum_{r=1}^{\infty} \hat{\rho}_l^{-1} \langle \hat{\xi}_j, \Psi_{22}(\hat{\phi}_r) \rangle \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tr} \hat{\pi}_{tl} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \\
=& \sum_{l=1}^{q_W} \hat{\rho}_l^{-1} \langle \hat{\xi}_j, \Psi_{22}(\hat{\phi}_l) \rangle \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl}^2 \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \\
=& \sum_{l=1}^{q_W} \langle \hat{\xi}_j, \Psi_{22}(\hat{\phi}_l) \rangle \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \quad . \qquad (127)
\end{aligned}$$

Similarly, for the first part we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \langle \hat{\xi}_j, \Psi_{22}(W_t) \rangle Z_t' = \sum_{l=1}^{\infty} \langle \hat{\xi}_j, \Psi_{22}(\hat{\phi}_l) \rangle \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \qquad (128)$$

Summing over q_f terms, the difference between the two parts is

$$\begin{aligned}
&\left\| \sum_{j=1}^{q_f} \sum_{l=q_W+1}^{\infty} \langle \hat{\xi}_j, \Psi_{22}(\hat{\phi}_l) \rangle \hat{\xi}_j \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \right\| \\
=& \left\| \sum_{j=1}^{q_f} \langle \hat{\xi}_j, \sum_{l=q_W+1}^{\infty} \Psi_{22}(\hat{\phi}_l) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \rangle \hat{\xi}_j \right\| \\
\xrightarrow{p}& \left\| \sum_{j=1}^{q_f} \langle \hat{\xi}_j, \sum_{l=q_W+1}^{\infty} \Psi_{22}(\hat{\phi}_l) \iota_l \rangle \hat{\xi}_j \right\| \ . \qquad (129)
\end{aligned}$$

where ι_l is $o_p(1)$ following the argument in Lemma 6 (Step 2). Moreover, by Assumption 4(ii) Ψ_{22} is Hilbert-Schmidt and thus $\Psi_{22}(\hat{\phi}_l)$ is bounded. Therefore, the above term converges to zero.

Term 3.

Consider again the *j-th* term, we end up with

$$\langle \hat{\xi}_j, \frac{1}{T} \sum_{t=1}^{T} (\hat{f}_t - f_t) Z_t' \rangle - \sum_{l=1}^{q_W} \hat{\rho}_l^{-1} \left(\langle \hat{\xi}_j, \frac{1}{T} \sum_{t=1}^{T} (\hat{f}_t - f_t) \hat{\pi}_{tl} \rangle \right) \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\pi}_{tl} Z_t' \right) \ . \qquad (130)$$

By Lemma 4, we have

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - f_t)Z_t' = o_p(1), \quad \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - f_t)\hat{\pi}_{tl} = o_p(1) \tag{131}$$

Term 4.

Following Lemma 6, we have

$$\frac{1}{T}\sum_{j=1}^{q_f}\hat{\xi}_j\langle\hat{\xi}_j, e_f\rangle\hat{M}_\pi Z' = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{q_f}\hat{\xi}_j\langle\hat{\xi}_j, e_{ft}\rangle\right)\left[z_t' - \sum_{l=1}^{q_W}\frac{\langle\hat{W}_t,\hat{\phi}_l\rangle\hat{\Gamma}_{WZ}(\hat{\phi}_l)'}{\hat{\rho}_l}\right]$$
$$= \left(\sum_{j=1}^{q_f}\hat{\xi}_j\langle\hat{\xi}_j, \frac{1}{T}\sum_{t=1}^{T}e_{ft}z_t'\rangle\right) - \sum_{j=1}^{q_f}\sum_{l=1}^{q_W}\hat{\rho}_l^{-1}\hat{\xi}_j\frac{\sum_{t=1}^{T}\langle\hat{\xi}_j, e_{ft}\rangle\langle\hat{W}_t,\hat{\phi}_l\rangle}{T}\hat{\Gamma}_{WZ}(\hat{\phi}_l)'$$

By Assumption 4(ii), e_{ft} is an iid H-white noise. Hence the first term converges to zero in L^2 as

$$\|Ee_{ft}z_t'\| = 0\ . \tag{132}$$

Similarly, the second term converges to zero as

$$\frac{\sum_{t=1}^{T}\langle\hat{\xi}_j, e_{ft}\rangle\langle\hat{W}_t,\hat{\phi}_l\rangle}{T} = \iint\underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}e_{ft}(u)\hat{W}_t(v)\right)}_{\xrightarrow{p}0\text{ in }L^2}\hat{\xi}_j(u)\hat{\phi}_l(v)dudv\ . \tag{133}$$

■

Proof. **(Lemma 9)**

Given Lemma 7 and 8, the proof of $\hat{\Psi}_{22}$ is trivial. By the decomposition of $\hat{\eta}$ in (118), we have

$$\hat{\Psi}_{22} = \left[\langle\hat{\xi}, \Psi_{22}(\hat{W})\rangle + \langle\hat{\xi}, \Psi_{22}(W-\hat{W})\rangle + \langle\hat{\xi}, e_f\rangle + \langle\hat{\xi}, \hat{f} - f\rangle\right]M_Z\hat{\pi}'\hat{\Omega}_{\pi Z\pi}. \tag{134}$$

By Lemma 7, the term with $\langle\hat{\xi}, \Psi_{22}(W-\hat{W})\rangle$ goes to zero; by Lemma 8, the terms with $\langle\hat{\xi}, e_f\rangle$ and $\langle\hat{\xi}, \hat{f} - f\rangle$ converge to zero. Hence, multiplying them by $\hat{\xi}_j(u)\hat{\phi}_k(v)$ yields mean zero random functions. Given that, we can simply focus on the term with $\langle\hat{\xi}, \Psi_{22}(\hat{W})\rangle$.

Notice that (similar to Lemma 7),

$$\Psi_{22}(\hat{W}) = \Psi_{22}(\hat{\phi})\hat{\pi} + \left[\sum_{l=q_W+1}^{\infty} \hat{\pi}_{1,l}\Psi_{22}(\hat{\phi}_l), \ldots, \sum_{l=q_W+1}^{\infty} \hat{\pi}_{T,l}\Psi_{22}(\hat{\phi}_l)\right] . \tag{135}$$

We can write down the estimated kernel $\hat{\psi}_{22}$ by

$$\begin{aligned}&\sum_{i=1}^{q_f}\sum_{j=1}^{q_W}\langle\hat{\xi}_i, \Psi_{22}(\hat{\phi}_j)\rangle\hat{\xi}_i(u)\hat{\phi}_j(v)-\\&\sum_{i=1}^{q_f}\sum_{j=1}^{q_W}\frac{\sum_{t=1}^{T}\sum_{l=q_W+1}^{\infty}\langle\hat{\xi}_i, \Psi_{22}(\hat{\phi}_l)\rangle\hat{\pi}_{tl}}{T}\left[\hat{\pi}_t' - Z_t'\hat{\Gamma}_Z\hat{\Gamma}_{WZ}(\hat{\phi})\right]\hat{\Omega}_{\pi Z\pi}^{-1}[i,j]\hat{\xi}_i(u)\hat{\phi}_j(v)\end{aligned} \tag{136}$$

. The second term converges to mean zero random functions following exactly the same argument as before. The difference between the first term and the true kernel is

$$\begin{aligned}&\|\sum_{i=1}^{q_f}\sum_{j=1}^{q_W}\langle\hat{\xi}_i, \Psi_{22}(\hat{\phi}_j)\rangle\hat{\xi}_i(u)\hat{\phi}_j(v) - \psi_{22}\|\\=&\|\sum_{i=1}^{q_f}\sum_{j=1}^{q_W}\langle\hat{\xi}_i, \langle\psi_{22}, \hat{\phi}_j\rangle\rangle\hat{\xi}_i(u)\hat{\phi}_j(v) - \psi_{22}\|\end{aligned} \tag{137}$$

By Assumption 3, both q_f and q_W go to infinity and by Assumption 4(ii) the operator Ψ_{22} is bounded. Hence the above term converges to zero. ■

Appendix B Data Preprocessing

First, I discuss the construction of household income data. Specifically, net household income before the deduction of housing costs (ENTINCHH) is first computed by adding net earnings including self-employed income, net investment income, net occupational pensions, benefits, children income and other income sources, and subtracting deductions. Before deducting the housing costs, net household income (BHC) is adjusted using the Survey of Personal Incomes (SPI) that contains information of the very rich households. In particular, rich households in the samples are re-weighted to reflect the total number of the very rich in the UK. After adjusted using SPI data, the net household income after the housing costs (ESAHCHH) is computed by subtracting the housing costs. Further, ESAHCHH is divided by OECD equivalence scales, and the results should be interpreted as equivalized income for a childless couple. Finally, the income is deflated using the associated price series. Details of the construction of HBAI is provided in Goodman and Webb (1994).

As for the SVAR step, the aggregate variables are obtained from Cloyne (2013). Moreover, the 90-10 ratio is computed using the percentiles weighted by the sampling frequency. As is shown in Figure 10, the percentiles computed using the HBAI dataset closely match the yearly Living Standards Poverty and Inequality (LSPI) data constructed by the IFS (Keiller et al., 2020). The mean log deviations (MLD) is computed by

$$\text{MLD} = ln(\bar{x}_t) - \bar{ln x_{it}}$$

where x_{it} are the income of household i at time t and $\bar{x}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} x_{it}$ is the mean income. By definition, observations with negative income are dropped.

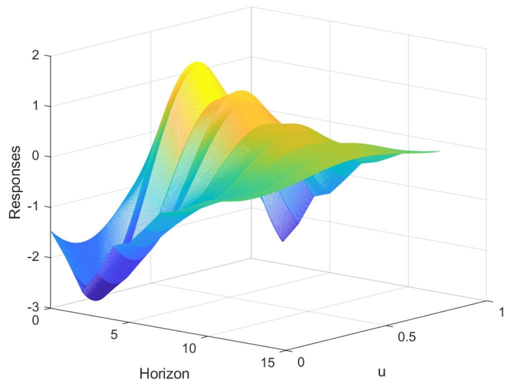
Tables and figures

Table 1: Simulation design: parameter values

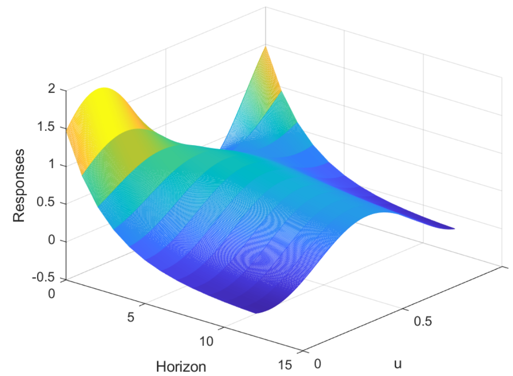
	Design 1	*Design 2*
a_{yy}	0.4	0.8
a_{yf}	(-0.3, 0.2, 0.1)	(-0.6, -0.2, -0.1)
a_{fy}	(-0.6, -0.3, -0.4)	(0.5, 0.3, 0.4)
a_{ff}	$\begin{bmatrix} -0.05 & -0.23 & 0.76 \\ 0.8 & -0.05 & 0.04 \\ 0.04 & 0.76 & 0.23 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.8 \end{bmatrix}$
σ_{11}	1	1
σ_{12}	(1, 1/2, 1/3)	(1, 1, 1)
σ_{21}	(0, 0, 0)	(0, 0, 0)
σ_{22}	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Table 2: Simulation Results

	Horizon	*Design 1* MISE	UE	Coverage	*Design 2* MISE	UE	Coverage
T=100	0	0.0366	0.2394	35.08	0.0464	0.2628	35.45
	1	0.0597	0.3375	85.33	0.0731	0.3400	86.09
	2	0.0581	0.3405	88.28	0.1099	0.4187	84.54
	3	0.0622	0.3612	89.00	0.1223	0.4471	91.45
	4	0.0548	0.3335	89.91	0.5722	1.1072	60.59
	5	0.0453	0.3053	90.34	2.1007	2.2932	24.51
	6	0.0420	0.2930	90.37	4.0710	3.2445	17.87
	7	0.0361	0.2700	91.17	4.6036	3.4572	16.78
	8	0.0296	0.2458	91.97	2.9225	2.7284	19.37
	9	0.0244	0.2235	92.91	0.7514	1.2805	54.25
	10	0.0192	0.1966	93.86	0.1872	0.5758	82.02
	11	0.0152	0.1708	94.57	0.8806	1.4271	42.83
	12	0.0119	0.1480	95.81	0.9236	1.4734	39.92
T=200	0	0.0157	0.1610	24.80	0.0188	0.1723	25.51
	1	0.0262	0.2291	86.49	0.0305	0.2255	85.93
	2	0.0265	0.2341	88.89	0.0463	0.2768	86.93
	3	0.0314	0.2594	87.65	0.0875	0.3947	86.67
	4	0.0268	0.2343	88.55	0.6577	1.2645	32.60
	5	0.0210	0.2095	89.80	2.3676	2.4846	17.13
	6	0.0192	0.1989	91.11	4.4934	3.4459	14.21
	7	0.0167	0.1836	92.28	5.0647	3.6617	13.66
	8	0.0147	0.1725	93.07	3.2699	2.9308	15.18
	9	0.0126	0.1593	94.03	0.8758	1.4700	26.44
	10	0.0105	0.1438	94.49	0.0811	0.3745	88.72
	11	0.0089	0.1305	94.57	0.6406	1.2368	35.40
	12	0.0071	0.1153	95.27	0.6969	1.2993	32.41
T=500	0	0.0057	0.0982	16.47	0.0062	0.1026	15.54
	1	0.0092	0.1365	87.47	0.0101	0.1315	86.94
	2	0.0097	0.1425	89.03	0.0158	0.1624	89.52
	3	0.0144	0.1766	84.16	0.0750	0.3962	68.72
	4	0.0113	0.1527	87.07	0.7088	1.3558	19.35
	5	0.0078	0.1289	88.26	2.5061	2.5798	13.36
	6	0.0072	0.1229	90.51	4.7012	3.5418	11.39
	7	0.0064	0.1152	92.87	5.2873	3.7572	10.98
	8	0.0065	0.1161	91.77	3.4389	3.0256	12.05
	9	0.0057	0.1077	92.26	0.9401	1.5636	16.92
	10	0.0050	0.1006	93.57	0.0318	0.2309	91.80
	11	0.0048	0.0968	92.51	0.5186	1.1444	21.41
	12	0.0038	0.0855	91.98	0.5753	1.2098	19.72

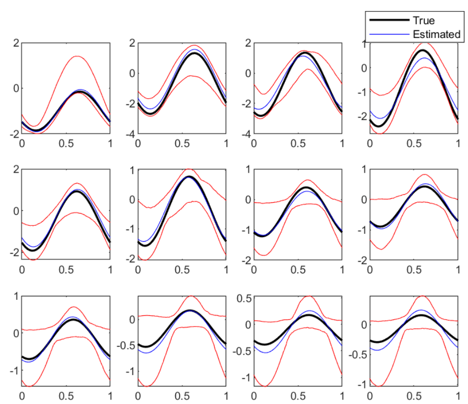


(A) Design 1

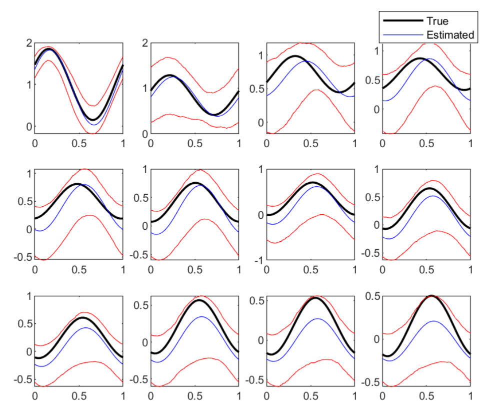


(B) Design 2

Figure 1: Simulation: True IR

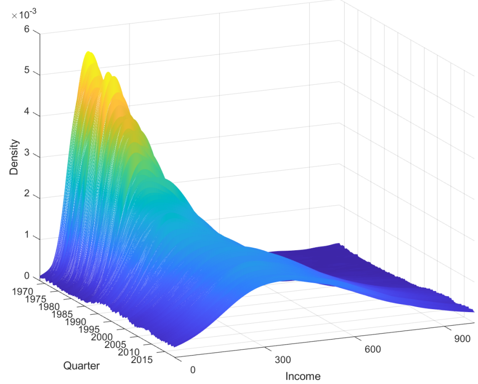


(A) Design 1

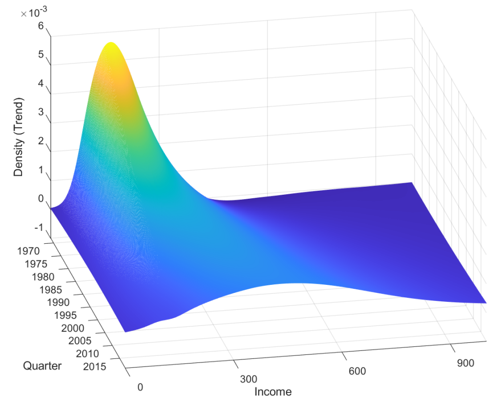


(B) Design 2

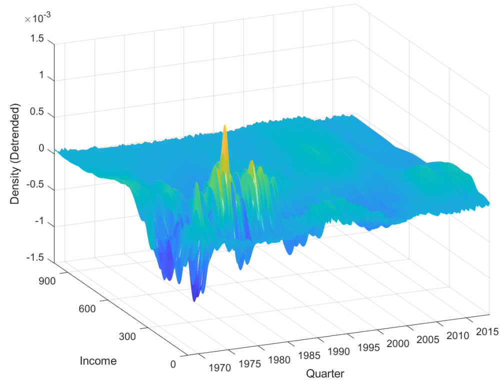
Figure 2: Simulation: Example



(A) Income Density



(B) Income Density (Trend)



(C) Income Density (Detrended)

Figure 3: Income Density

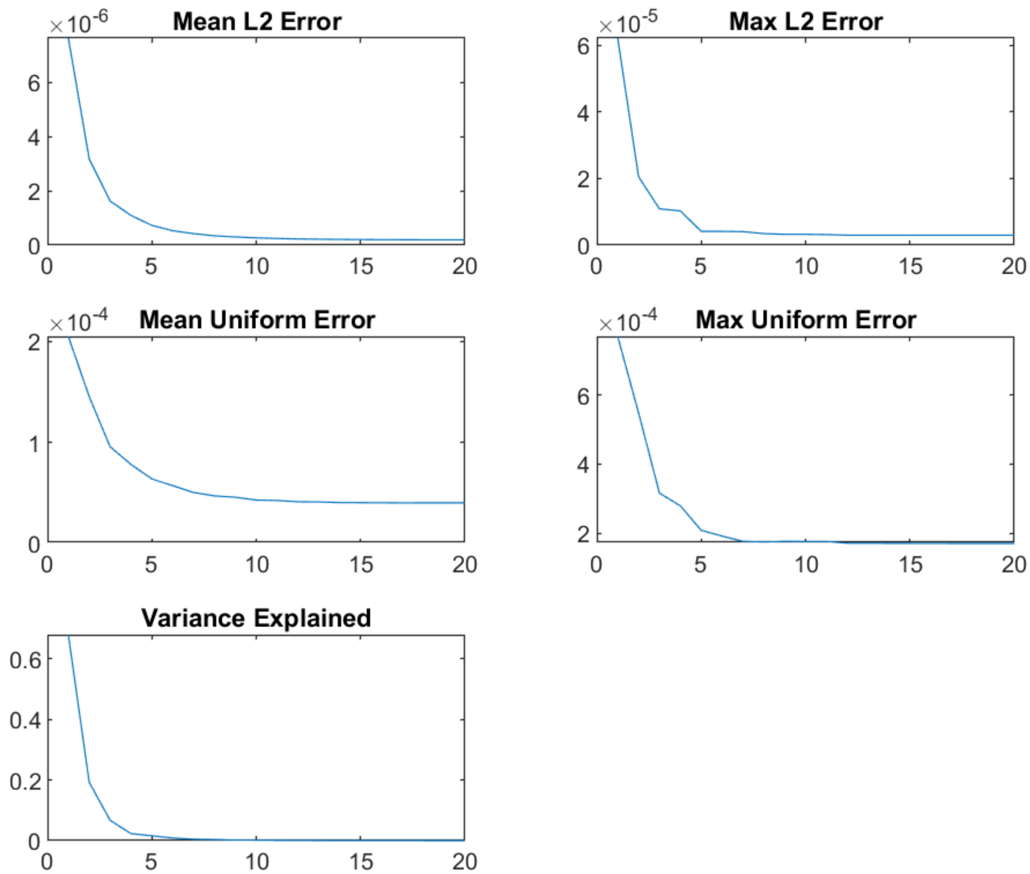


Figure 4: FPCA Criterion

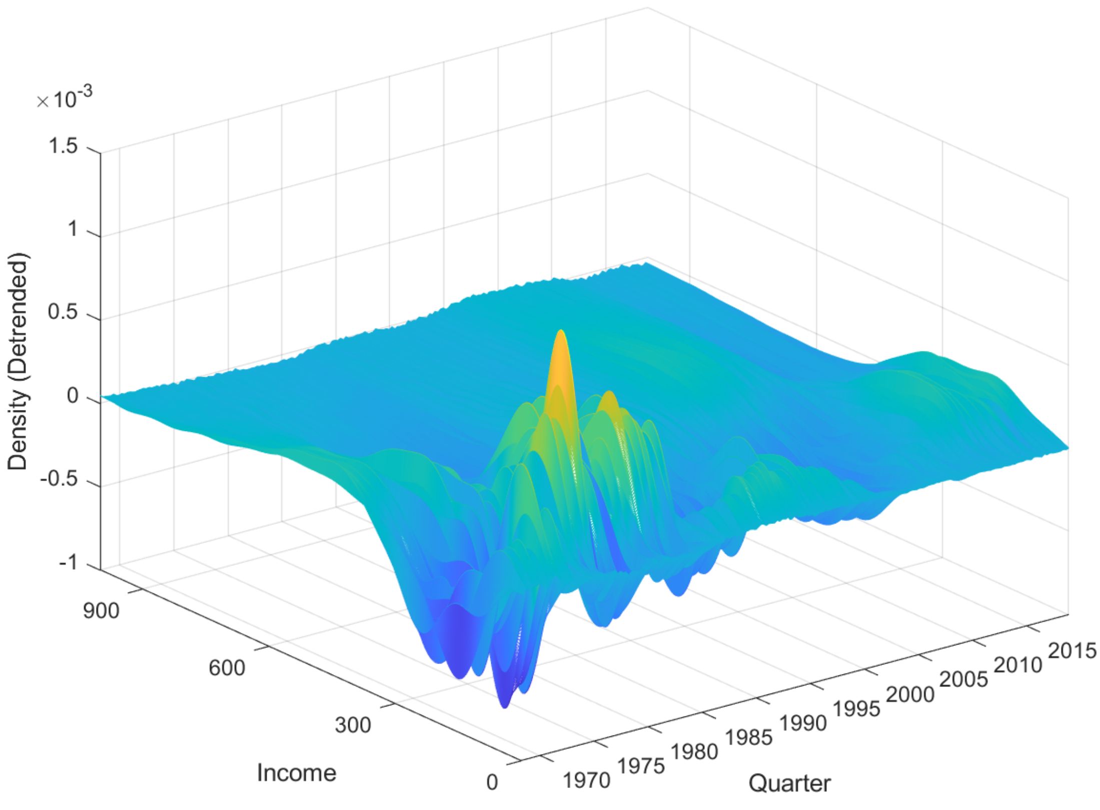


Figure 5: Density Recovered by FPCA

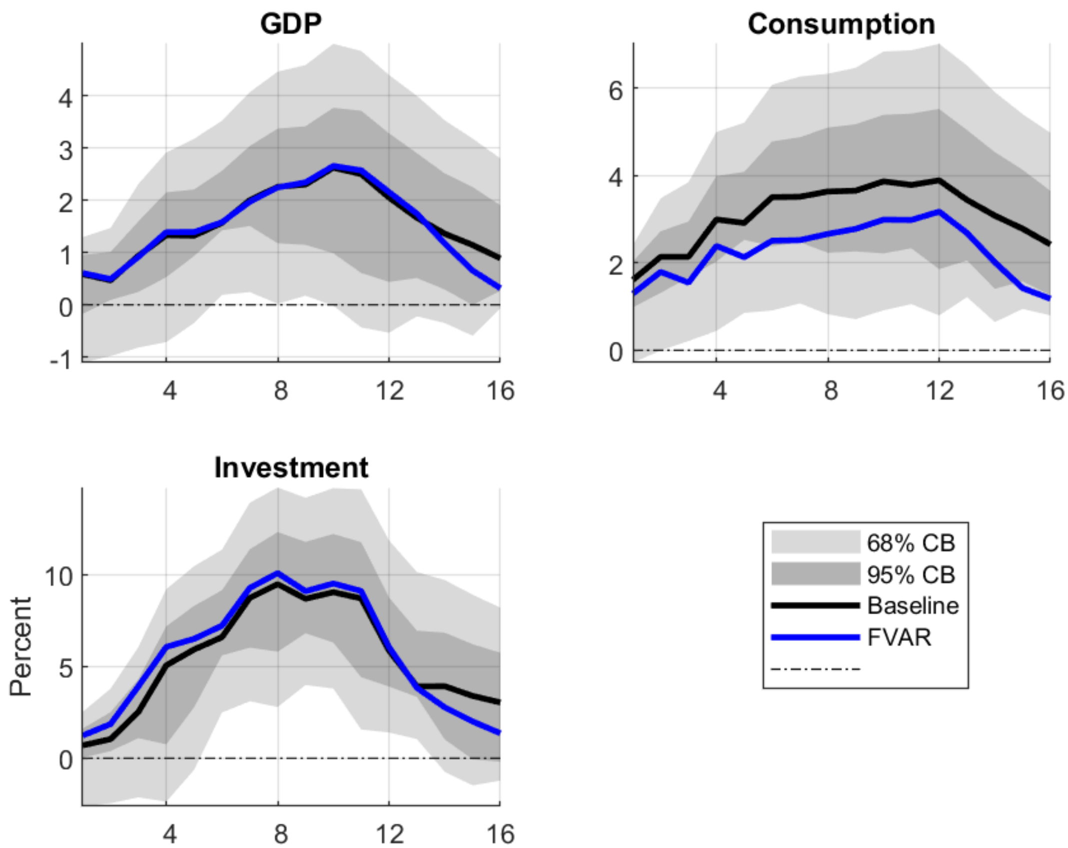
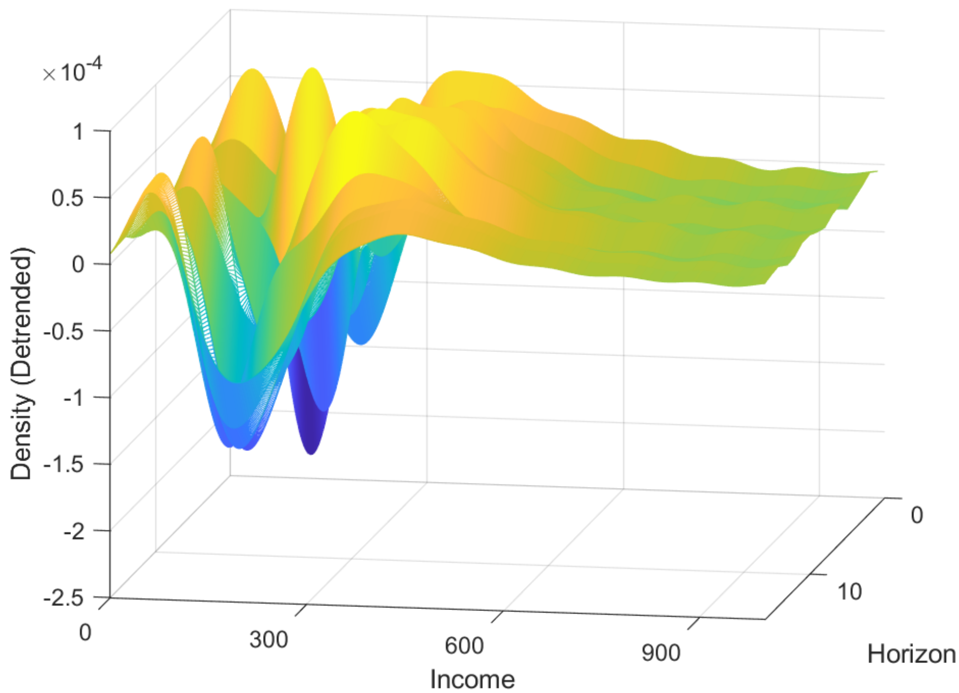
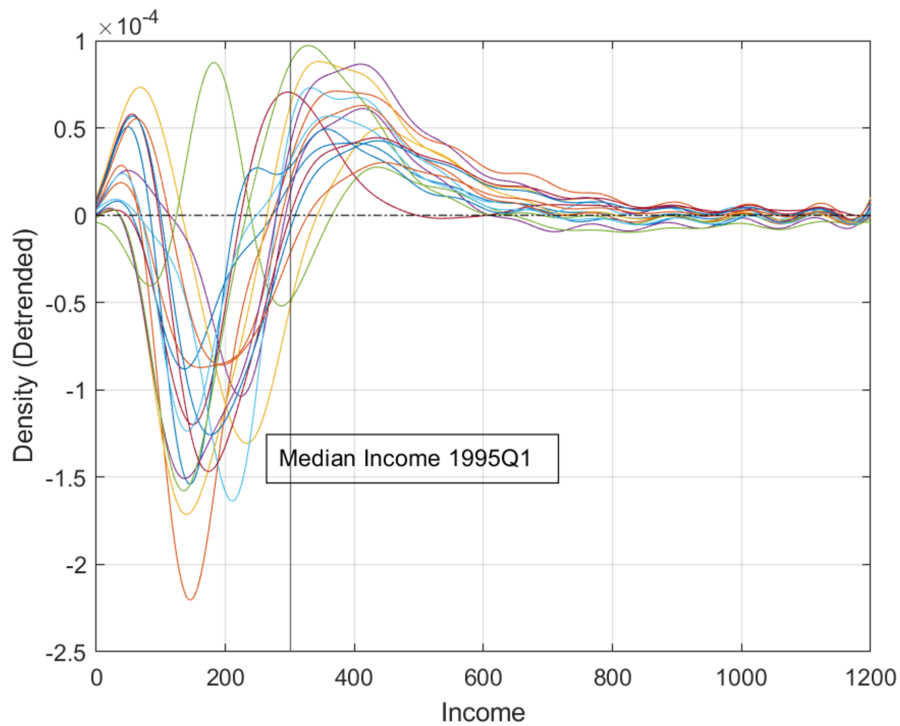


Figure 6: Responses of Aggregates

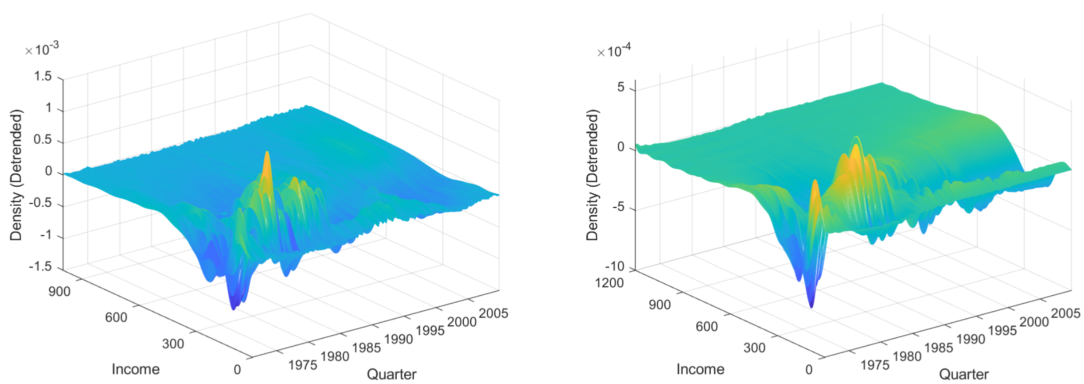


(a) Income Density Responses Over Horizon



(b) Income Density Responses (Stacked)

Figure 7: Distributional Effects of Tax Cuts



(A) Actual Density (B) Simulated Density

Figure 8: Simulated Income Density (Detrended)

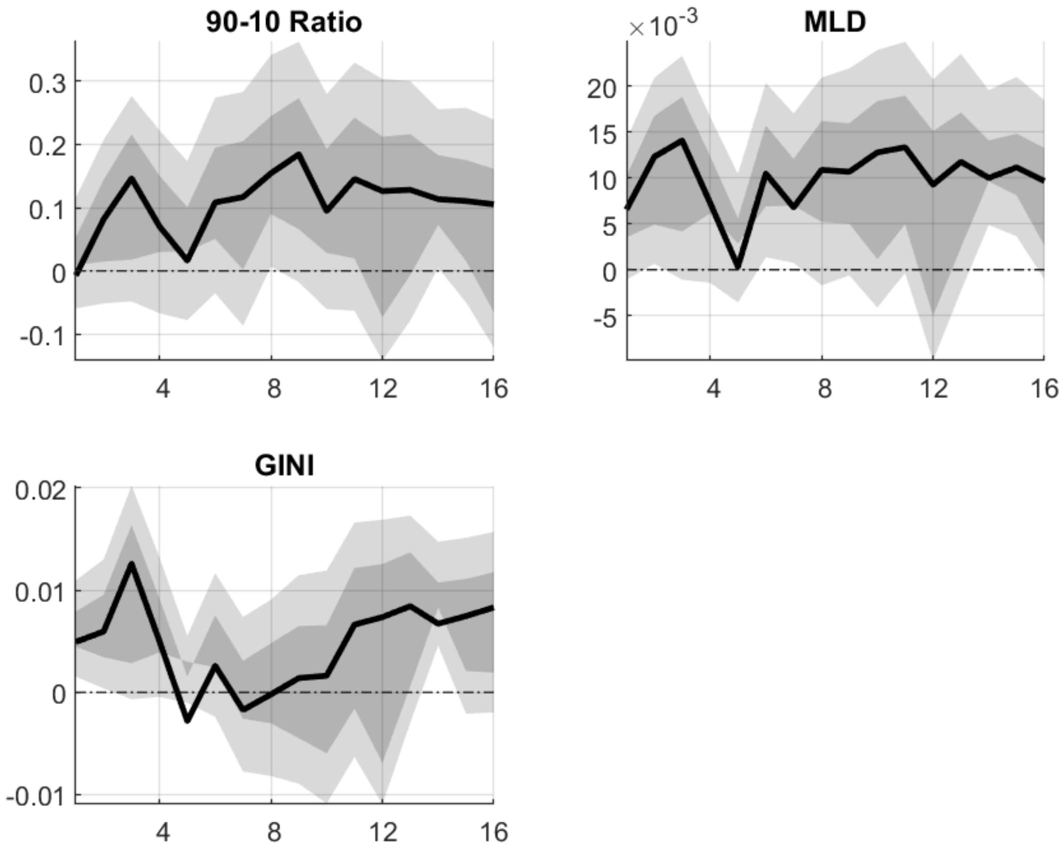


Figure 9: Responses of Inequality Measures

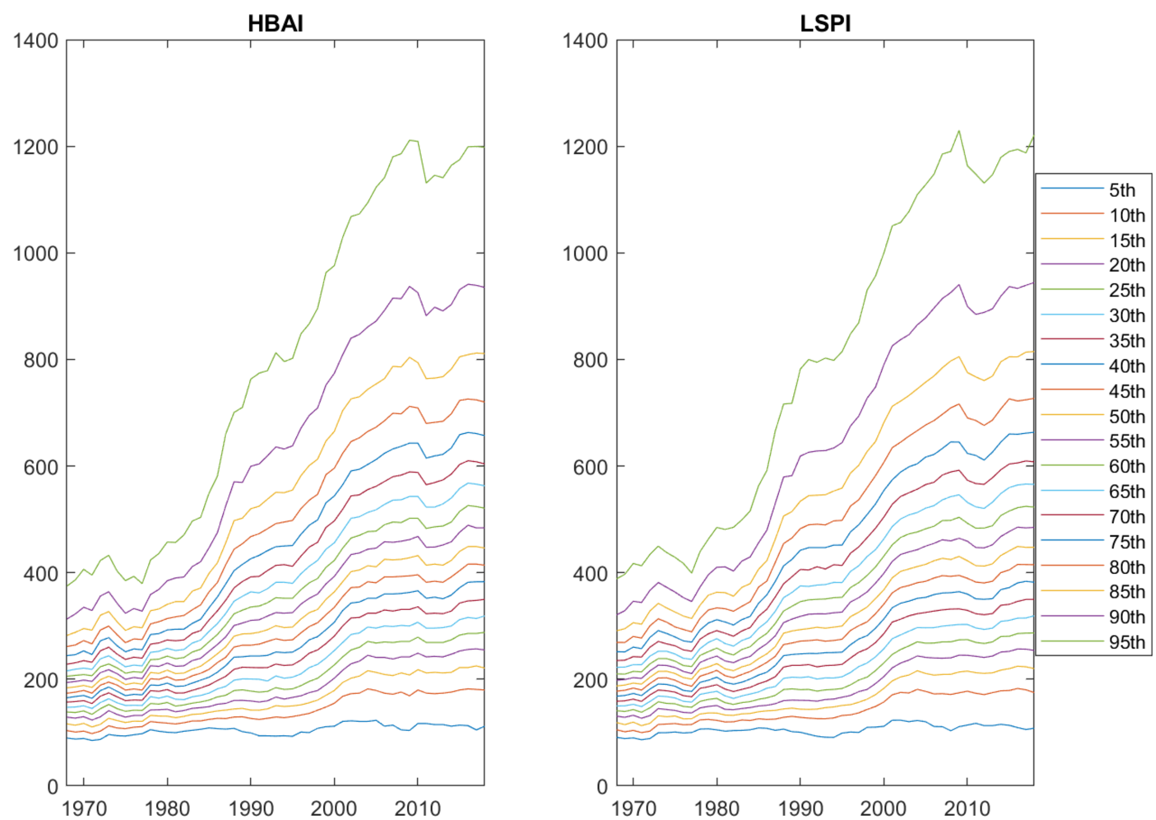


Figure 10: Comparison of Percentiles