Estimation and Application of Random Vector Fields: A Local Polynomial Regression Approach

Davide Fiaschi * Marco Francischello† Angela Parenti‡
Cristiano Ricci§

February 28, 2025

Abstract

This paper introduces a methodology for estimating a Random Vector Field (RVF), which characterises the local direction of movement for a sample of observational units in a d-dimensional space. We develop an estimator for RVF based on Local Polynomial Regression (LPR) and establish its asymptotic normality. Additionally, we propose a local significance test and introduce a methodology for optimal bandwidth selection. To address small sample bias, we explore the effectiveness of the Adaptive Kernel (AK) approach and inference via bootstrap. We illustrate the properties of our estimator by analysing the joint dynamics of GDP per capita and life expectancy — the so-called Preston Curve — for a sample of 105 countries from 1960 to 2015. Finally, we generate a forecast for 2045 based on the estimated RVF, demonstrating one of its key applications in economic and demographic analysis.

Keywords: nonparametric estimates, local polynomial regression, optimal bandwidth, adaptive kernel, Preston curve

JEL: C14; O51; R11

*Davide Fiaschi, University of Pisa,Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: davide.fiaschi@unipi.it).

†Marco Francischello, University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: marco.francischello@ec.unipi.it).

‡Angela Parenti (corresponding author), University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: angela.parenti@unipi.it).

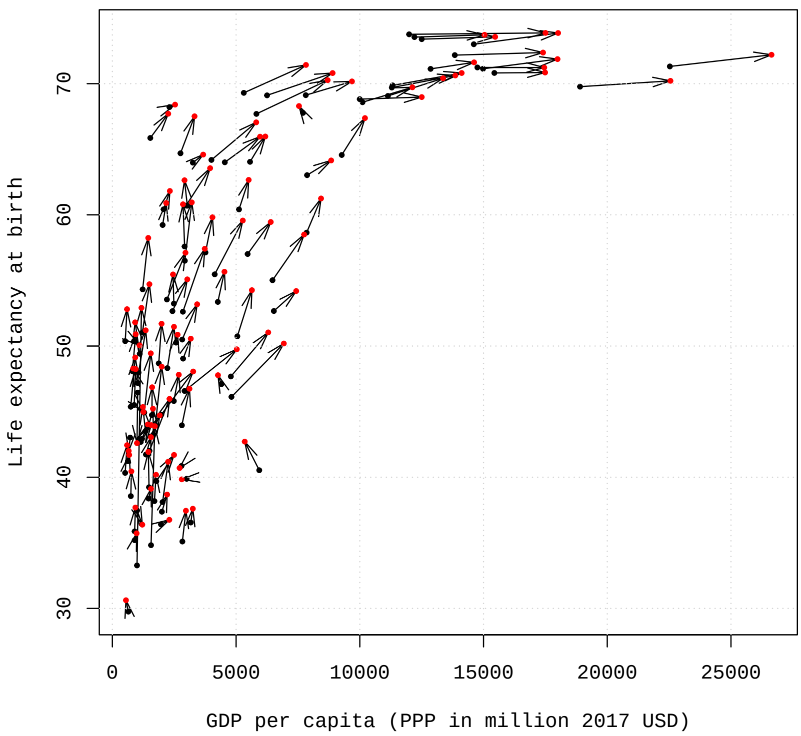
§Cristiano Ricci, University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: cristiano.ricci@unipi.it).

Contents

1 Introduction

Several economic phenomena exhibit heterogeneous local dynamics, meaning that the behaviour of observational units is crucially affected by their current status. For example, Figure 1 illustrates the heterogeneity of local dynamics in GDP per capita and life expectancy for a sample of 105 countries during the period 1960–1965 (Preston, 1975).

Figure 1: The joint dynamics of life expectancy at birth and real GDP per capita (PPP in million 2017 USD) during the period 1960–1965 for a sample of 105 countries (black (red) points represent observations in 1960 (2015)).



Source: Life expectancy at birth from the World Development Indicators; GDP per capita from the Penn World Table 10.01.

While the cross-sectional relationship between life expectancy at birth and income per capita, known as the Preston Curve, has been extensively studied in the literature (Bloom and Canning, 2007), its evolution over time—i.e., the dynamics shown in Figure 1—has received less attention (Easterlin, 2004), partly due to the absence of a robust econometric methodology that accounts for local dynamics.

This paper aims to address this gap—or more broadly, to estimate the local direction of movement for a sample of observational units in a d-dimensional space when such movement is modelled by a *Random Vector Field* (RVF). We propose an RVF estimator based on *Local Polynomial Regression* (LPR) and demonstrate that, under standard conditions, it is asymptotically normally distributed. Additionally, we introduce a testing procedure to assess the significance of local movements along specific directions. Furthermore, we develop a methodology for selecting the optimal bandwidth for LPR and explore how the *Adaptive Kernel* (AK) approach mitigates local bias in small sample cases. We illustrate our methodology by estimating the dynamics of the Preston Curve from 1960 to 2015 for a sample of 105 countries. Finally, we use a forecast for 2045 to showcase one of RVF's key potential applications.

From a methodological perspective, our paper relates to two distinct strands of literature. The first is nonparametric econometrics (Li and Racine, 2007), particularly *Local Polynomial Regression*, which we discuss as an efficient method for estimating an RVF (Györfi et al., 2013; Fan and Gijbels, 2018). We also draw on the VAR literature, as an RVF can be interpreted as a *nonparametric VAR* under certain specifications. In this regard, Härdle et al. (1998) provides the closest methodological contribution to our estimation approach, while Kalli and Griffin (2018) introduces a Bayesian method to mitigate the curse of dimensionality, a common challenge in nonparametric estimation.

The second strand of literature relates to the estimation of *conditional kernel density*, first introduced by Rosenblatt (1969). Among others, (Quah, 1992) pioneered its use in estimating the stochastic kernel to analyse the distribution dynamics of income per capita. The stochastic kernel can be viewed as a generalisation of our RVF in one dimension, where the error component of the RVF is not assumed to be additive. Fiaschi et al. (2018) provides an initial attempt to extend this approach to a two-dimensional space. However, while conditional kernel density offers greater generality, it also increases complexity in both estimation and inference.

In our empirical application, this paper contributes to the ongoing debate on the joint dynamics of life expectancy and income per capita—commonly known as the Preston Curve (Preston, 1975)—which carries significant policy implications (Pritchett, 2024). The literature on this topic has largely adopted a cross-sectional perspective (Bloom and Canning, 2007), with a few notable exceptions that take a more descriptive approach (Easterlin, 2004).

Our paper advances the literature in several key ways. First, we propose a robust econometric framework for estimating an RVF under general conditions, drawing on the extensive literature on Local Polynomial Regression. In this context, we establish a specific theorem for the limiting distribution of our RVF estimator, distinguishing our work from Li and Racine (2007) and Härdle et al. (1998). Second, we incorporate the adaptive kernel approach and develop a methodology for optimal bandwidth selection, significantly improving estimation quality in cases where distributions are not "well-behaved". Finally, our application to the Preston Curve underscores the importance of adopting a dynamic perspective on this topic. We identify key nonlinearities in the joint dynamics of life expectancy and GDP per capita, contributing to the ongoing debate on the subject (Fiaschi et al., 2020; Pritchett, 2024).

This paper is organised as follows. Section 2 introduces our approach to estimating an RVF, its limiting distribution, and additional details on estimation. Section 3 explores the numerical properties of our estimator. Section 4 applies our methodology to the dynamics of GDP per capita and life expectancy. Finally, Section 5 presents our conclusions.

2 Estimation of a random vector field

Section 2.1 introduces the theoretical framework of random vector field (RVF) in discrete time in a d-dimensional space. Section 2.2 introduce our estimator of RVF based on a local linear estimator and discuss its limiting behaviour. Section 2.3 presents a modified version of our estimator based on adaptive kernel, which is particularly useful in the case of non-uniform distribution of observations. Section 2.4 proposes a method for the choice of bandwidth. Finally, Section 2.5 contains some practical indications on how to run inference via bootstrap in the small sample case and, in general, when the asymptotic properties of the estimator seem to be likely to fail.

2.1 The econometric model

We consider a *vector field* $F : \mathbb{R}^d \to \mathbb{R}^d$ acting on a sample of a random variable $X_t \in \mathbb{R}^d$

$$\Delta X_{i,t} = F(X_{i,t-1}) + \Sigma(X_{i,t-1})^{1/2}\epsilon_{i,t} \quad i = 1, \dots, N, \tag{1}$$

where $\Delta X_{i,t} := X_{i,t} - X_{i,t-1}$ is the time change in $X_{i,t}$, and where, conditional on $X_{i,t-1}$ the $\epsilon_{i,t}$ are i.i.d. random variables with zero mean and identity covariance matrix. Equation (1) represents the time evolution of the observations X_{it}; therefore, we will consider estimators of F based on the sample $(\Delta X_{1,t}, X_{1,t-1}), \ldots, (\Delta X_{N,t}, X_{N,t-1})$. This evolution presents both a deterministic and a stochastic component. The deterministic part, $F(X_{i,t})$, is a function of the current state. The stochastic part, $\Sigma(X_{i,t-1})^{1/2}\epsilon_{i,t}$ is a random noise whose intensity can depend on the observation $X_{i,t-1}$. The formulation of Equation (1) is quite general and flexible and can be seen as a nonparametric extension of a vector autoregressive process (VAR). To see why, consider the case where $F(X_{i,t-1}) := AX_{i,t-1}$, where $A \in \mathbb{R}^{d\times d}$ and $\Sigma(X^i_{i,t-1})^{1/2} := \Omega \in \mathbb{R}^{d\times d}$. In this case, Equation (1) reads:

$$X_{it} = (I + A)X_{i,t-1} + \Omega\epsilon_{i,t} = MX_{i,t-1} + \Omega\epsilon_{i,t} \quad i = 1, \ldots, N, \tag{2}$$

by defining $M := I + A$; this is the usual expression of a VAR (Hamilton, 1994). Equation (1) can be also accommodated to account for the dependence of F on some exogenous variable Z_i. In particular we can define $\tilde{X}_{i,t} := (X_{i,t}, Z^i)$ and $\tilde{F}(X_{i,t}, Z_i) := (F(X_{i,t}), Z_i)$ and we are again within the scope of Equation (1).

Finally, while the theoretical properties that we derive below hold true for a generic covariance structure $\Sigma(X_{t-1})$, in practice it is useful to linearly transform the $X_{i,t-1}$ so that their covariance matrix is the identity matrix. This will be the procedure that we will use in the numerical simulations in Section 3 and in the application in Section 4.

2.2 The estimation

To introduce our estimator of RVF we will start by considering each component $j = 1, \ldots, d$ of $\Delta X_{i,t}$ separately. More formally, we are interested in estimating F_j:

$$\Delta X_{i,t,j} = F_j(X_{i,t-1}) + \left[\Sigma(X_{i,t-1})^{1/2}\right]_j \epsilon_{i,t} \quad i = 1, \ldots, N, \tag{3}$$

where $\left[\Sigma(X_{i,t-1})^{1/2}\right]_j$ is the j-th row of the matrix $\Sigma(X_{i,t-1})^{1/2}$. Now, we discuss how the *Local Polynomial Regression* (LPR) can be used to estimate F_j. LPR is a *nonparametric* technique that consists in locally approximating F in a neighbourhood of the evaluation point x with a polynomial in the X_i. The estimator is local in the sense that through the use of kernel functions, observations close to x contribute more to the value of the estimator in x. We focus on the *local constant* estimator, also known as *Nadaraya-Watson* estimator, and the *local linear* estimator. In particular, the Nadaraya-Watson estimator is defined as:

$$\widehat{F}_{0,j}(x) := \sum_{i=1}^{N} \frac{K\left(\frac{X_{i,t-1}-x}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{X_{i,t-1}-x}{h}\right)} \Delta X_{i,t,j}, \tag{4}$$

where h is the *bandwidth*, $K : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$K(x) = k(x_1) \cdots k(x_d), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{5}$$

and $k : \mathbb{R} \to \mathbb{R}$ is a bounded non-negative kernel function, that satisfies

$$\int_{\mathbb{R}} k(x)\, dx = 1, \quad k(x) = k(-x), \quad \int_{\mathbb{R}} x^2 k(x)\, dx = k_2 < +\infty, \quad \int_{\mathbb{R}} k(x)^2\, dx = ||k||_2^2 < +\infty. \tag{6}$$

It can be shown that (Li and Racine, 2007, pp. 63):

$$\widehat{F}_{0,j}(x) = \hat{a}_{0,j}(x) = \arg\min_{a \in \mathbb{R}} \sum_{i=1}^{N} \left(\Delta X_{i,t,j} - a\right)^2 K\left(\frac{X_{i,t-1} - x}{h}\right). \tag{7}$$

Similarly, we can define the local linear estimator as follows. Consider the minimizer $(\hat{a}_{1,j}(x), \hat{b}_{1,j}(x)^T) \in \mathbb{R}^{d+1}$ of:

$$(\hat{a}_{1,j}(x), \hat{b}_{1,j}(x)^T) := \arg \min_{a \in \mathbb{R}, b \in \mathbb{R}^d} \sum_{i=1}^{N} \left[\Delta X_{i,t,j} - \left(a + b^T X_{i,t-1}\right)\right]^2 K\left(\frac{X_{i,t-1} - x}{h}\right); \tag{8}$$

then the local linear estimator of F_j is given by:

$$\widehat{F}_{1,j}(x) := \hat{a}_{1,j}(x). \tag{9}$$

It can also be shown that $\hat{b}_{1,j}(x)$ is an estimator of the gradient of F_j at x (Li and Racine, 2007, pp. 82). From Equations (4) and (8) we can verify how we locally approximate F with a polynomial in $X_{i,t-1}$ and how the kernel K weights the contribution of each observation i proportionally to the distance between $X_{i,t-1}$ and x.

Here, we will limit our attention only to local constant and local linear estimators; however, but LPR can be defined for any degree p (Li and Racine, 2007).The level of computational complexity of the estimator grows with the degree of the polynomial but also the accuracy of the estimator. In practice it is rare to use polynomials of degree higher than 3, and, usually, the selection falls on polynomial estimators with odd degree (say, 1 and 3) because they can be shown to have the same variance than the even degree ones that precedes them but smaller bias (Fan and Gijbels, 2018). Furthermore, as we will see in the next section, for $p \in \{0, 1\}$, a local polynomial estimator of degree p has no bias while estimating a vector field whose $p + 1$-th derivatives are null (Fan and Gijbels, 2018).

2.2.1 Limit distribution

The two theorems below report the limiting distribution

Theorem 2.1 (Asymptotic normality of component-wise estimator). *Let K be a kernel satisfying Eqq* (5) *and* (6). *Furthermore assume that the j-th component of the vector field F, $F_j(x)$, the density of $X_{i,t-1}$, $f(x)$, and the covariance matrix of the errors, $\Sigma(x)$, are three times differentiable. As $h \to 0$, $N \to \infty$, $Nh^{d+2} \to \infty$ and $Nh^{d+6} \to 0$, for any interior point x for the local constant estimator it holds*

$$(Nh^d)^{1/2} \left[\widehat{F}_{0,j}(x) - F_j(x) - \frac{\kappa_2 h^2}{2} \sum_{s=1}^{d} \left(\frac{2}{f(x)} \frac{\partial f}{\partial x_s} \frac{\partial F_j}{\partial x_s}(x) + \frac{\partial^2 F_j}{\partial x_s^2}(x) \right) \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{||k||_2^{2d}\, \Sigma_{j,j}(x)}{f(x)} \right), \tag{10}$$

while for the local linear estimator

$$(Nh^d)^{1/2} \left[\widehat{F}_{1,j}(x) - F_j(x) - \frac{\kappa_2 h^2}{2} \sum_{s=1}^{d} \frac{\partial^2 F_j}{\partial x_s^2}(x) \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{||k||_2^{2d}\, \Sigma_{j,j}(x)}{f(x)} \right), \tag{11}$$

where $\mathcal{N}$ is the Normal distribution.

Proof. See Chapter 2 of Li and Racine (2007). □

Theorem 2.1 can be interpreted as a description of the limit distribution of the marginals of the vectors $\widehat{F}_0(x) = \left(\widehat{F}_{j,0}\right)_{j=1,\ldots,d}$ and $\widehat{F}_1(x) = \left(\widehat{F}_{j,0}\right)_{j=1,\ldots,d}$. Since we are interested in estimating RVF, we want to understand also the limit distribution of the whole estimator and not only its marginals. To do this we can extend the results of the previous theorem:

Theorem 2.2 (Asymptotic normality of the RVF estimator)**.** *Under the same assumptions of Theorem 2.1 it holds that for the joint distribution of $\widehat{F}_0(x)$ and $\widehat{F}_1(x)$ we have:*

$$(Nh^d)^{1/2}\left[\widehat{F}_0(x) - F(x) - \frac{\kappa_2 h^2}{2}\left[\frac{2}{f(x)}\mathcal{J}F(x)\nabla f(x) + \Delta F(x)\right]\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{||k||_2^{2d}\,\Sigma(x)}{f(x)}\right), \tag{12}$$

and

$$(Nh^d)^{1/2}\left[\widehat{F}_1(x) - F(x) - \frac{\kappa_2 h^2}{2}\Delta F(x)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{||k||_2^{2d}\,\Sigma(x)}{f(x)}\right), \tag{13}$$

where $\mathcal{J}F(x)$ is the Jacobian matrix of $F(x)$, $\nabla f(x)$ is the gradient of $f(x)$, and $\Delta F(x)$ is the vector Laplacian of $F(x)$, i.e.

$$\mathcal{J}F(x) \in \mathbb{R}^{d\times d}, \quad \mathcal{J}F(x)_{r,s} = \frac{\partial F_r(x)}{\partial x_s}, \tag{14}$$

$$\nabla f(x) \in \mathbb{R}^d, \quad \nabla f(x)_r = \frac{\partial f(x)}{\partial x_r}, \quad \Delta F(x) \in \mathbb{R}^d, \quad \Delta F(x)_r = \sum_{s=1}^{d} \frac{\partial^2 F_r(x)}{\partial x_s^2}. \tag{15}$$

Proof. The proof follows the steps of Theorem 2.1, with the additional step that now we have to compute the covariance terms between any two components of the vector estimator. This computation can be carried out by using the same set of techniques used in computing the variance in Theorem 2.1. □

Finally, Theorem 2.2 makes clear why, as we mentioned above, a local linear estimator has no bias if F is a local linear RVF, i.e. $\Delta F(x) = 0\ \forall x$.

2.3 Estimation with the adaptive kernel

We also propose an adaptive kernel version of our estimator. In the same flavour adaptive kernel density estimation (Silverman, 1986), by considering a local bandwidth that varies with the local density of initial observations. The adaptive approach is particularly effective when observations are spread very far apart from the median, with regions displaying a high number of observations and others with only a few. We then replace Equations (8) and (9) by

$$(\hat{a}_{1,j}^{ad}(x), \hat{b_{1,j}^{T,ad}}(x)) := \arg\min_{a\in\mathbb{R}, b\in\mathbb{R}^d} \sum_{i=1}^{N} \left(\Delta X_{i,t,j} - \left(a + b^T X_{i,t-1}\right)\right)^2 K\left(\frac{X_{i,t-1} - x}{h_i}\right). \tag{16}$$

and

$$\widehat{F}_{1,j}^{ad}(x) := \hat{a_{1,j}^{ad}}(x). \tag{17}$$

In practice, the choice of h_i is carried out by performing first a pilot estimation of the density of observations at time t,

$$\hat{f}_0(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h^d}K\left(\frac{X_{i,t-1}-x}{h}\right), \tag{18}$$

and then choosing $h_i = h \cdot \lambda_i$ with

$$\lambda_i = \left(\frac{\hat{f}_0(X_{i,t-1})}{g}\right)^{-\alpha} \text{ and } g = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(\hat{f}_0(X_{i,t-1}))\right). \tag{19}$$

Parameter $\alpha \in (0,1)$, measuring the intensity of adaptiveness, i.e. how the estimator is sensible to the variation of local density, is usually taken equal to 0.5 as a reference value. A possible approach in the selection of parameter α (as well as h) is described in Section 2.4.

2.4 Optimal choice of bandwidth

The choice of parameters α and h of Equations (16) and (17) are made by minimising the mean square error (MSE) of the estimate for a set of candidate couples. In particular, for each possible couple of (α, h) estimate the adaptive RVF, $\widehat{F}_1^{ad}(x; h, \alpha)$; then, use the estimated RVF, given the initial observations, to predict their values in the final year, i.e.

$$\widehat{X}_{i,t} := X_{i,t-1} + \widehat{F}_1^{ad}(X_{i,t-1}; h, \alpha), \tag{20}$$

Finally, the goodness of the estimation is measured by the mean of the square distance between the predicted observations $\widehat{X}_{i,t}$ and the actual observation at the final year $X_{i,t}$, i.e.:

$$MSE(\alpha, h) := \frac{1}{N}\sum_{j=1}^{N}\left|\left|X_{t-1} + \widehat{F}_1^{ad}(X_{t-1}; h, \alpha) - X_t^i\right|\right|^2.$$

In practice, the optimal couple of (α, h), i.e. the one minimising $MSE(\alpha, h)$, is searched over a finite grid.

2.5 Inference

Theorem 2.2 allows us to address the simplest scenario in the estimation process, i.e., the case with a large number of observations with "well-behaved" distributional properties. In particular, an estimated direction at x is statistically significant if we can reject the null hypothesis that all d components of movement are jointly equal to zero. Therefore, given the estimated direction at x, $\widehat{F}_1(x)$, and the estimated asymptotic covariance matrix of Theorem 2.2, it is possible to show that under the null hypothesis:

$$\widehat{F}_1(x)' \left[\frac{||k||_2^{2d} \, \widehat{\Sigma}(x)}{\hat{f}(x) N h^d} \right]^{-1} \widehat{F}_1(x) \sim \chi_d^2, \tag{21}$$

which represents a standard Wald test, whose limiting distribution can be found in Greene (2003, p. 107).

Other scenarios, for example, the use of the adaptive kernel, small sample case, not well-behaved distribution of observations, cross-sectional dependence and autocorrelated observations, pose significant issues for the inference. In this case, we can resort to a bootstrap procedure, with the caveat that sampling should be performed from the set of all observed pairs $(\Delta X_{i,t}, X_{i,t-1})$. Additionally, a version of the bootstrap method (e.g., block bootstrap) should be adopted to mitigate potential issues arising from the sample's inherent properties, and observations with cross-sectional and time dependence.

3 Numerical simulations

In this section, we numerically explore the properties of proposed LPR estimator, using a VAR representation as a benchmark (Equation (2)). As discussed above, the latter represents a sort of "linear" representation of an RVF. In particular, we consider a 2-dimensional space and two data generation processes, the first where the linearity of VAR should have an advantage in the estimate (Section 3.1). Instead, in the second case, the nonlinear dynamics should make evident the advantages of using a local linear estimator (Section 3.2). In all simulations, we use as an initial condition a sample of 10,000 observations drawn from a bivariate Gaussian distribution with mean $(0, 0)$ and diagonal covariance matrix and constant variance of 0.75. RVF is estimated by the local linear estimator (i.e. LPR of order 1) using the Epanechnikov kernel and the optimal choice of bandwidth. The VAR is estimated as suggested Hamilton (1994) by OLS.

3.1 Linear RVF

Consider the following RVF:

$$\Delta X_{i,t} = M X_{i,t-1} + \Omega \epsilon_{i,t}, \text{ for } i = 1, \ldots, N, \tag{22}$$

where $X_{i,t} = (X_{i,t,1}, X_{i,t,2})$ is a vector of length 2, ϵ_t is a vector of length 2 of independent Gaussian errors with unit variance,

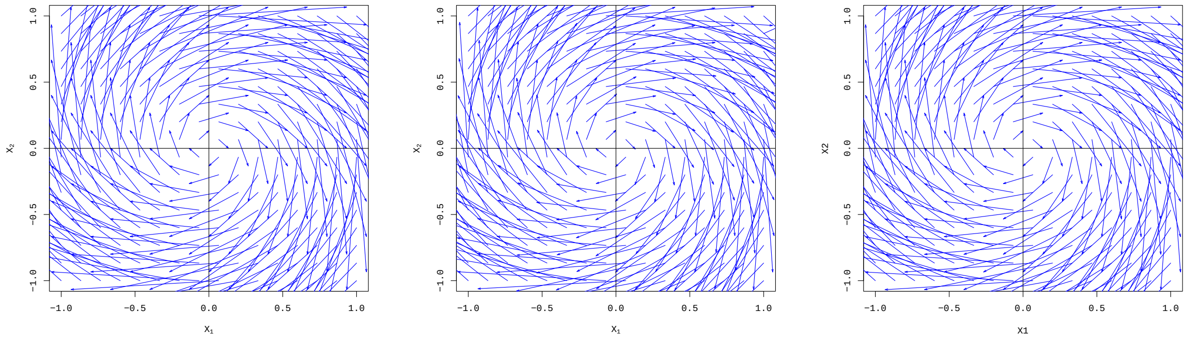
$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{23}$$

and

$$\Omega = \begin{bmatrix} 0.01 & 0.005 \\ 0.005 & 0.02 \end{bmatrix}. \tag{24}$$

The specification of M makes the direction of vectors circular (the eigenvalues have null real parts and complex conjugates), as shown in Figure 2a. The focus on a linear RVF as the one in Equation (22) is motivated by the goal of comparing local linear estimation with the one resulting from using a VAR representation, where the latter should perform better because data generation process and econometric specification perfectly coincide. Comparing Figures 2b and 2c with Figure 2a, we can appreciate how both methods identify the overall shape of RVF. However, a closer inspection of (log of) local errors calculated as $\log_{10}\left(\left|\left|\widehat{F}(x) - F(x)\right|\right| / ||F(x)||\right)$ ($\widehat{F}(x)$ is the estimated RVF in x and $F(x)$ its true value), reported in Figure 3, reveals the better performance of VAR concerning the local linear estimator, with the local error mainly in the range $0.01\% - 10\%$ (10^{-3}–10^{-1}) for the local linear estimator versus 0.001%–0.01% (10^{-4}–10^{-3}) for the VAR. The worst performances for both estimates are in the origin, where the direction of RVF is particularly involved. The use of adaptive kernel as explained in Section 2.3 (with the standard value of $\alpha = 0.5$) does not produce substantial improvements in the estimate but a higher precision around the origin, where the error is higher. The higher density of observations around the origin, indeed, produces a lower bandwidth there

Figure 2: Local linear model versus VAR representation in the estimate of a linear RVF generated using Equation (22), with the specification of M in (23) and of the covariance matrix Ω in (24) for a sample of 10,000 observations.



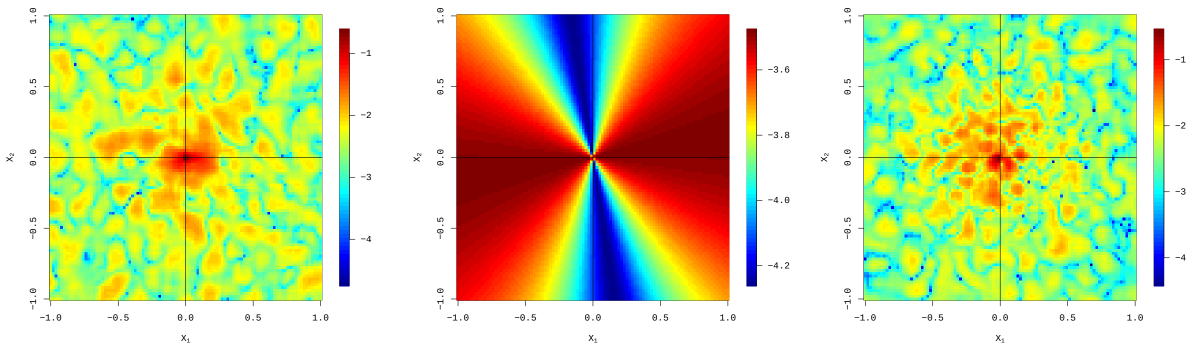
(a) The true RVF (periodic pattern).

(b) The estimate of a linear RVF via local linear estimator. Only vectors statistically significant at 5% level are reported (tests are made using Theorem 2.2).

(c) The estimate of linear RVF by a VAR representation. Only vectors statistically significant at 5% level are reported.

and therefore, in the presence of very local dynamics, a higher precision.

Figure 3: The local errors in the estimation of linear RVF calculated as $\log_{10}\left(\left|\left|\widehat{F}(x) - F(x)\right|\right| / ||F(x)||\right)$ ($\widehat{F}(x)$ is the estimated RVF in x and $F(x)$ is its true value).



(a) The estimated local errors for the local linear estimator.

(b) The estimated local errors for the VAR representation.

(c) The estimated local errors for the local linear estimator using the adaptive kernel of Section 2.3.

3.2 Nonlinear RVF

Consider the following RVF:

$$\Delta X_t := X_t - X_{t-1} = F(X_{t-1}) + \Omega \epsilon_t, \tag{25}$$

where:

$$F(X_{i,t}) = -\frac{1}{5} \begin{bmatrix} 2X_{i,t,1}^3 - X_{i,t,1} \\ X_{i,t,2} \end{bmatrix} \tag{26}$$

and Ω is the same of (24). Figure 4a highlights the presence of three attractors, all on the x-axis, where the ones at the extreme are locally stable, while the one in the origin is unstable (in physics, this pattern is called the symmetric double-well potential). The extreme nonlinearity is well managed by the local linear estimator, as shown in Figure 4b, which reproduces the overall pattern. This does not hold for the VAR representation, which turns out to be incapable of grasping even the distinctive features of the aggregate dynamics (Figure 4c). The estimated local errors fully reflect these conclusions, with the error for the local linear estimator mainly ranging in $0.01\% - 30\%$ ($10^{-3} - 10^{-0.5}$) versus a main range of $0.01\% - 10000\%$ ($10^{-3} - 10^{1}$). The use of adaptive kernel again does not produce sensible improvements in the estimate but a more uniform distribution of errors (see Figure 5c).

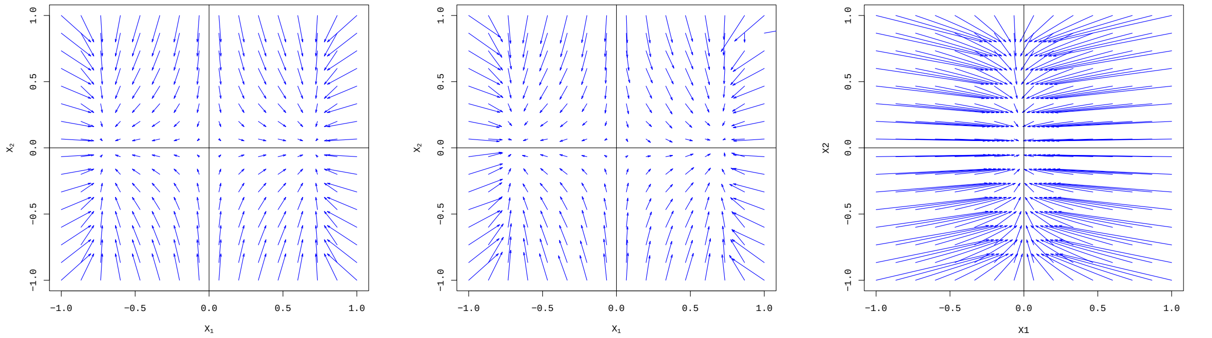
4 The dynamics of Preston Curve

Figure 6 presents the estimated relationship between the logarithm of GDP per capita and life expectancy at birth at country level for the years 1960-2015, based on a sample of 105 countries using local constant estimator (i.e. LPR of order 0).[1] The Preston Curve is the term used in the literature to describe this positive relationship, which assumes a characteristic concave shape in its original version, where GDP per capita is not taken in logarithmic scale (Preston, 1975).

Comparing the curves for 1960 (blue curve) and 2015 (red curve) in Figure 6, the Preston Curve appears to shift northwest over time. As explained by Easterlin (2004), this shift results from two (possibly interconnected) processes, which are useful to con-

[1] Data on life expectancy is taken from World Development Indicators Bank (2025), while on GDP per capita from Penn World Table 10.01 Feenstra et al. (2021).

Figure 4: Local linear model versus VAR representation in the estimate of a nonlinear RVF generated using Equation (25), with the specification of F in (26) and of the covariance matrix Ω in (24) for a sample of 10,000 observations.

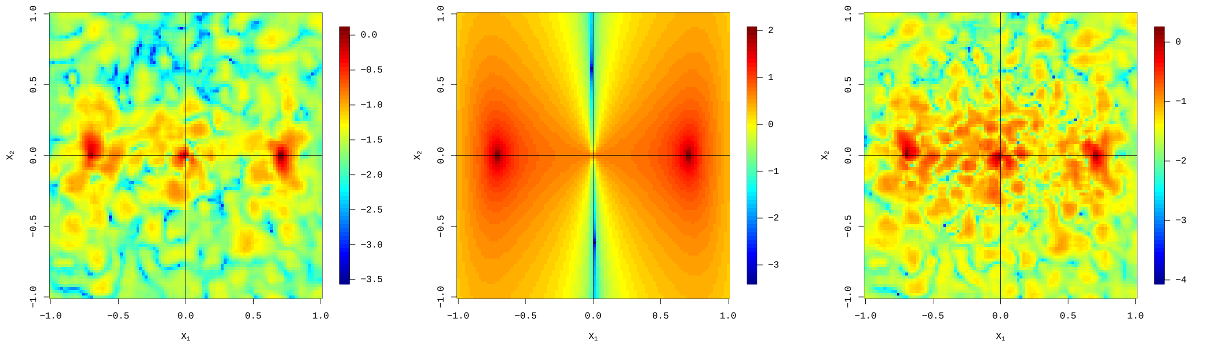


(a) The true RVF (multiple equilibria pattern).

(b) The estimate of nonlinear RVF via local linear estimator. Only vectors statistically significant at 5% level are reported (tests are made using Theorem 2.2).

(c) The estimate of linear RVF by a VAR representation. Only vectors statistically significant at 5% level are reported.

Figure 5: The local errors in the estimation of nonlinear RVF calculated as $\log_{10}\left(\left|\left|\widehat{F}(x) - F(x)\right|\right| / ||F(x)||\right)$ ($\widehat{F}(x)$ is the estimated RVF in x and $F(x)$ is its true value).



(a) The estimated local errors for the local linear estimator.

(b) The estimated local errors for the VAR representation.

(c) The estimated local errors for the local linear estimator with the adaptive bandwidth.

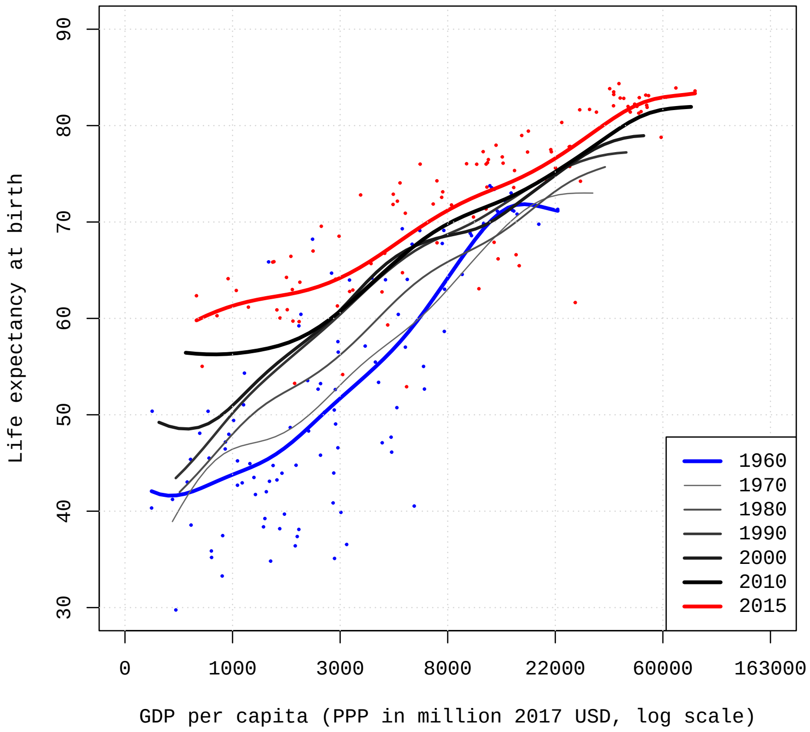
sider separately. The first is the increase in life expectancy due to the development and introduction of improved medical technologies and practices, reflected in the upward shift of the curve—i.e., for a given level of income, life expectancy increases over time. The second is the rise in per capita GDP due to factor accumulation and technological progress, reflected in the horizontal shift of the curve—i.e., for a given level of life expectancy.

Our focus on the dynamics of the Preston Curve also suggests considering the logarithm of GDP per capita instead of its absolute level, as is done in the original Preston Curve. GDP per capita generally exhibits exponential growth over time, and using its absolute difference over time in the estimation would involve a non-stationary variable. In contrast, taking the logarithm of GDP per capita ensures that horizontal movements in the regression function correspond to GDP per capita growth rates, which should be time-stationary if some form of conditional convergence across countries' incomes has held over the period of analysis (Barro and Sala-i Martin, 2004).

In all the estimates we use the local linear estimator (i.e. the LPR of order 1), an Epanechnikov kernel and the optimal choice of bandwidth. The significance level of estimated RVF is calculated using the results in Theorem 2.2. Figure 7a presents the estimated RVF for our sample of 105 countries, considering 5-year changes (resulting in a total sample size of 1,070 observations). Using a grid of $100 \times 100 = 10,000$ evaluation points, the reported arrows indicate the estimated local directions in the RVF that are statistically significant at the 5% level. The overall pattern suggests a joint increase in both variables, but with important nuances that can be better appreciated by also examining Figure 7b. For low levels of life expectancy and GDP per capita, changes in the former are more pronounced compared to those in the latter. This is particularly evident for countries with a life expectancy of 40 years and a GDP per capita of $3000. Beyond 70 years of life expectancy, this gradient remains but becomes progressively weaker.

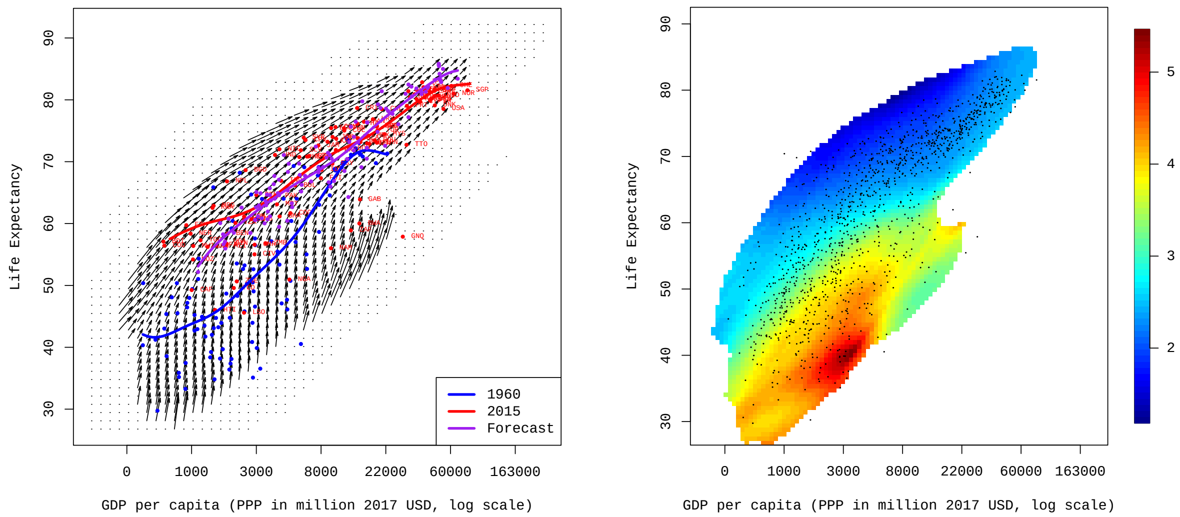
Figure 8a highlights a key advantage of our methodology—namely, the ability to provide a local estimate of the direction of the RVF. In particular, we can observe that, for a given level of life expectancy, the growth rate of GDP per capita decreases as GDP per capita increases (i.e., evidence of convergence), except within the life expectancy range of [45,60], where the relationship follows a U-shaped pattern. The magnitude of

Figure 6: The estimated relationship, obtained using LPR of order 0, between the logarithm of GDP per capita and life expectancy at birth for various years in a sample of 105 countries. Blue (red) points represent observations from 1960 (2015).



Source: Life expectancy at birth from the World Development Indicators and GDP per capita from the Penn World Table 10.01.

Figure 7: The estimated RVF for a sample of 105 countries over the period 1960–2015, using 5-year changes in life expectancy at birth and the logarithm of real GDP per capita (PPP in millions of 2017 USD).



(a) The arrows indicate the estimated local direction in the RVF. Only significant directions (at the 5% significance level) are reported (small black points are evaluation points where no significant direction is found). Blue (red) points represent observations from 1960 (2015). The Preston curves for 1960 and 2015 are estimated using LPR of order zero, while the forecast for 2015 is calculated using the procedure described in Section 2.4, starting from 1960.

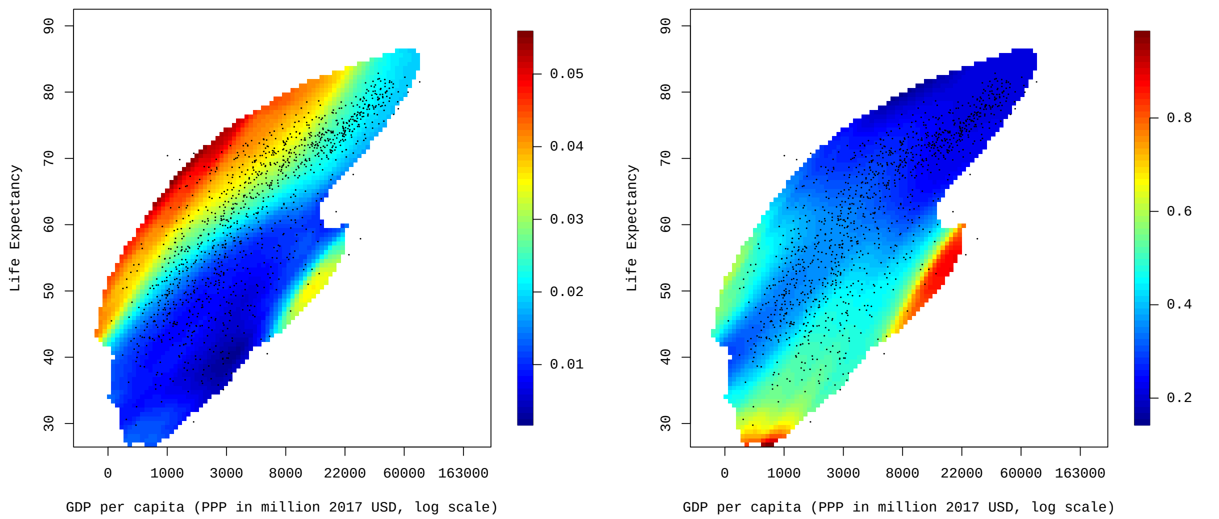
(b) The ratio of the change in life expectancy to the change in the logarithm of GDP per capita (i.e., the growth rate of GDP per capita). Small black points are the observations used in the estimate.

Source: Our calculations, using life expectancy at birth from the World Development Indicators and GDP per capita from the Penn World Table 10.01.

this gradient is also noteworthy, ranging from 0% to 5% in terms of the annual growth rate. Furthermore, for a given level of GDP per capita, countries with higher life expectancy tend to grow at a significantly faster rate, supporting the idea of a conditional convergence/growth regime model (Barro and Sala-i Martin, 2004; Easterlin, 2004; Fiaschi et al., 2020).

Equally interesting are the results of the estimated local changes in life expectancy, as reported in Figure 8b. For life expectancy levels above 70, no clear gradient is observed for either life expectancy or GDP per capita, suggesting a lack of convergence in life expectancy across countries. Below this threshold, however, for life expectancy in the range [45,60], changes in life expectancy exhibit a U-shaped relationship with GDP per capita—indicating that the highest increases in life expectancy have occurred in both poor and middle-income countries. A similar pattern emerges when considering GDP per capita: for countries with a GDP per capita below $ 1,000, the changes in life expectancy display a comparable U-shaped trend. This evidence further suggests that the trajectories followed by countries with poor health conditions are complex and cannot be adequately explained by a single theoretical framework (Easterlin, 2004; Bloom and Canning, 2007).

Figure 8: The separate estimate of the two components of the vectors in the RVF of Preston Curve. Small black points are the observations used in the estimate.



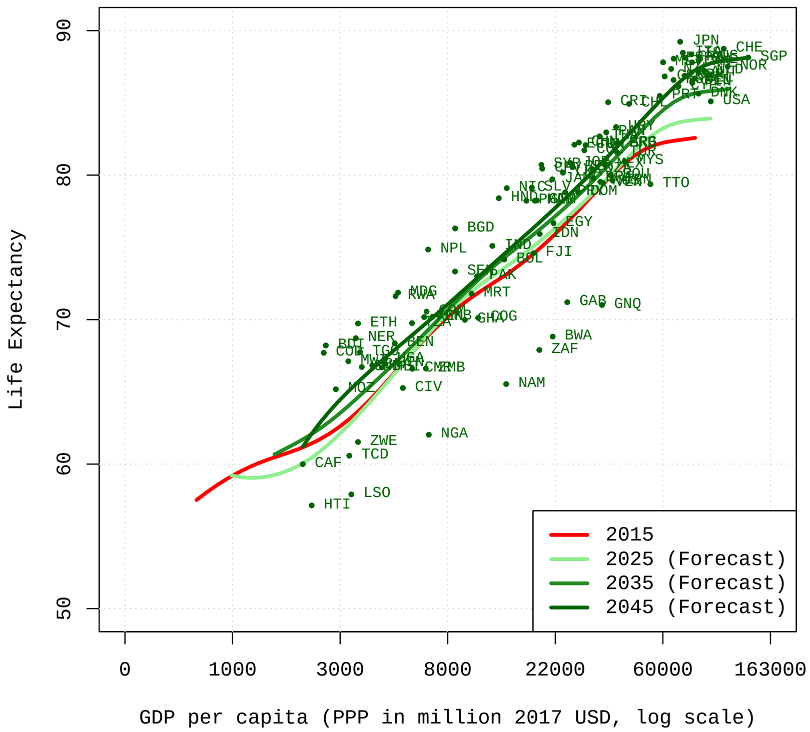
(a) The estimated direction of the logarithm of GDP per capita scaled to annual changes (annual growth rate of GDP per capita).

(b) The estimated direction of life expectancy scaled to annual change (absolute change in life expectancy).

Source: our calculations on World Development Indicators and Penn World Table 10.01 data.

Finally, Figure 9 presents the estimated future dynamics of the Preston Curve based on the projected trends for our sample of countries, as forecasted using the estimated RVF, for the years 2025, 2035, and 2045. From Figure 9, the impression is that the phase of significant gains in life expectancy due to medical innovations is now over for low and medium countries, and only a substantial increase in their income can allow to achieve considerable improvements in their life expectancy. For high-income countries, instead, medical innovation or the adoption of best practices can still allow substantial gains (Easterlin, 2004). The longest-living country in our forecast of 2045 is Japan (also the one in 2015), which is expected to gain an average of 2.4 months per year in the next 30 years.

Figure 9: The 10-, 20-, and 30-year ahead forecast of the Preston Curve using the estimated RVF starting from 2015. Dark green points are the forecast in 2045 for our sample of 105 countries.



Source: our calculations on World Development Indicators and Penn World Table 10.01 data.

5 Conclusions

We have presented a methodology for estimating the RVF and discussed its limiting distribution. Numerical simulations have highlighted the importance of considering a nonparametric specification in the analysis of dynamical systems with strong local behaviour. The application to the Preston Curve has demonstrated the potential of our methodology in uncovering nonlinearities in thedynamics, as well as its usefulness for forecasting.

The paper leaves open several research questions, including: i) the limiting distribution of LPR when the adaptive kernel is used, a topic that has so far been studied only in density estimation and for specific estimators (Silverman, 1984); ii) the case where the RVF includes exogenous variables and higher-order lags, as in the parametric VAR literature (Hamilton, 1994); iii) the extension of our estimator to the three-dimensional case when the dynamics is confined to a sphere, which has important applications in spatial economics and other scientific fields (Marinucci and Peccati, 2011).

References

Bank, W. (2025). World development indicators. `https://databank.worldbank.org/source/world-development-indicators`.

Barro, R. and Sala-i Martin, X. (2004). *Economic growth*. Cambridge: The MIT Press.

Bloom, D. E. and Canning, D. (2007). Commentary: The preston curve 30 years on: still sparking fires. *International Journal of Epidemiology*, 36(3):498–499.

Easterlin, R. A. (2004). *The reluctant economist: Perspectives on economics, economic history, and demography*. Cambridge University Press.

Fan, J. and Gijbels, I. (2018). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Routledge.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2021). Penn world table 10.01. `https://www.rug.nl/ggdc/productivity/pwt/`.

Fiaschi, D., Gianmoena, L., and Parenti, A. (2018). Spatial club dynamics in European regions. *Regional Science and Urban Economics*, 72:115–130.

Fiaschi, D., Lavezzi, A. M., and Parenti, A. (2020). Deep and proximate determinants of the world income distribution. *Review of Income and Wealth*, 66(3):677–710.

Greene, W. H. (2003). *Econometric Analysis*. Pearson, New York, 5th edition.

Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (2013). *Nonparametric curve estimation from time series*, volume 60. Springer.

Hamilton, J. D. (1994). *Time series analysis*. Princeton university press.

Härdle, W., Tsybakov, A., and Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, 68(2):221–245.

Kalli, M. and Griffin, J. E. (2018). Bayesian nonparametric vector autoregressive models. *Journal of econometrics*, 203(2):267–282.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Marinucci, D. and Peccati, G. (2011). *Random fields on the sphere: representation, limit theorems and cosmological applications*, volume 389. Cambridge University Press.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population studies*, 29(2):231–248.

Pritchett, L. (2024). The case for economic growth as the path to better human wellbeing. *Economics*.

Quah, D. (1992). Empirical cross-section dynamics in economic growth. *European Economic Review*.

Rosenblatt, M. (1969). Conditional probability density and regression estimators. *Multivariate analysis II*, 25:31.

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *The annals of Statistics*, pages 898–916.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Appendix

A Countries'list

Code	Country	Code	Country	Code	Country
ARG	Argentina	GIN	Guinea	PER	Peru
AUS	Australia	GMB	Gambia	PHL	Philippines
AUT	Austria	GNB	Guinea-Bissau	PRT	Portugal
BDI	Burundi	GNQ	Equatorial Guinea	PRY	Paraguay
BEL	Belgium	GRC	Greece	ROU	Romania
BEN	Benin	GTM	Guatemala	RWA	Rwanda
BFA	Burkina Faso	HND	Honduras	SEN	Senegal
BGD	Bangladesh	HTI	Haiti	SGP	Singapore
BOL	Bolivia	IDN	Indonesia	SLV	El Salvador
BRA	Brazil	IND	India	SWE	Sweden
BRB	Barbados	IRN	Iran, Islamic Republic of	SYR	Syrian Arab Republic
BWA	Botswana	ISL	Iceland	TCD	Chad
CAF	Central African Republic	ITA	Italy	TGO	Togo
CAN	Canada	JAM	Jamaica	THA	Thailand
CHE	Switzerland	JOR	Jordan	TTO	Trinidad and Tobago
CHL	Chile	JPN	Japan	TUN	Tunisia
CHN	China	KEN	Kenya	TUR	Türkiye
CIV	Côte d'Ivoire	KOR	Korea, Republic of	TZA	Tanzania
CMR	Cameroon	LKA	Sri Lanka	UGA	Uganda
COD	Congo, DR	LSO	Lesotho	URY	Uruguay
COG	Congo	MAR	Morocco	USA	United States of America
COL	Colombia	MDG	Madagascar	VEN	Venezuela
COM	Comoros	MEX	Mexico	ZAF	South Africa
CPV	Cabo Verde	MLI	Mali	ZMB	Zambia
CRI	Costa Rica	MLT	Malta	ZWE	Zimbabwe
CYP	Cyprus	MOZ	Mozambique		
DEU	Germany	MRT	Mauritania		
DNK	Denmark	MUS	Mauritius		
DOM	Dominican Republic	MWI	Malawi		
DZA	Algeria	MYS	Malaysia		
ECU	Ecuador	NAM	Namibia		
EGY	Egypt	NER	Niger		
ESP	Spain	NGA	Nigeria		
ETH	Ethiopia	NIC	Nicaragua		
FIN	Finland	NLD	Netherlands		
FJI	Fiji	NOR	Norway		
FRA	France	NPL	Nepal		
GAB	Gabon	NZL	New Zealand		
GBR	Great Britain	PAK	Pakistan		
GHA	Ghana	PAN	Panama		