UMEÅ UNIVERSITY

Studying Earnings Trajectories as Functional Outcomes

Kreske Ecker

Department of Statistics
Umeå School of Business, Economics and Statistics
Umeå 2024

Doctoral Thesis
Department of Statistics
Umeå School of Business, Economics and Statistics
Umeå University
SE-901 87 Umeå

Copyright © 2024 by Kreske Ecker
Statistical Studies No. 58
ISBN: 978-91-8070-234-8 (print)
ISBN: 978-91-8070-235-5 (pdf)
ISSN: 1100-8989
Cover illustration by: Sen Nordstrom
Electronic version available at http://umu.diva-portal.org/

Printed by: Cityprint i Norr AB
Umeå, Sweden 2024

Contents

List of papers

The thesis is based on the following papers:

I. Ecker, K., de Luna, X. and Westerlund, O. (2022). Regional differences in initial labour market conditions and dynamics in lifetime income trajectories. *Longitudinal and Life Course Studies*, 13(3): 352-379.

II. Ecker, K., de Luna, X. and Schelin, L. (2023). Causal inference with a functional outcome. *Journal of the Royal Statistical Society, Series C: Applied Statistics.* In print.

III. Ecker, K., Baranowska-Rataj, A., Brydsten, A. and de Luna, X. (2023). The effects of initial firm age on earnings trajectories. *Manuscript.*

IV. Ecker, K., Liu, X., Schelin, L. and de Luna, X. (2023). Double robust estimation of functional outcomes with data missing at random. *Manuscript.*

Paper I is reprinted with the kind permission of Bristol University Press.
Paper II is reprinted with the kind permission of Oxford Academic.

Abstract

In this thesis, we present methods for studying patterns of income accumulation over time using functional data analysis. This is made possible by the availability of large-scale longitudinal register data in Sweden. By modelling individuals' cumulative earnings trajectories as continuous functions of time, we can explore temporal dynamics as well as divergences in these trajectories based on initial labour market conditions. A major contribution of this thesis consists of extending the potential outcome framework for causal inference to functional data analysis.

In Paper I, we use functional-on-scalar linear regression and an intervalwise testing procedure to study the associations between initial labour market size and income trajectories for one Swedish birth cohort. In Paper II, we present methods to draw causal conclusions in this setting. We introduce the functional average treatment effect (FATE), as well as an outcome-regression based estimator for this parameter. In addition, we show the finite sample distribution of this estimator under certain regularity conditions and demonstrate how simultaneous confidence bands can be used for inferences about the FATE. An application study in this paper estimates the causal effect of initial labour market size on income accumulation trajectories.

In Paper III, these methods are applied to study the effect of initial firm age on earnings accumulation. Paper IV presents an outcome regression based and a double robust estimator for the mean of functional outcomes when some of these outcome functions are missing at random. We derive the asymptotic distributions of these two estimators as well as their covariance structure under more general conditions.

Keywords: functional data analysis, causal inference, earnings trajectories, simultaneous confidence bands, missing data.

Populärvetenskaplig sammanfattning

I Sverige har vi en lång tradition av att samla in information om befolkningen för administrativa ändamål (till exempel beskattning) och att spara dessa uppgifter i register. Många av dessa uppgifter kan sedan även användas i forskningssyften. Till exempel har vi data om invånarnas årliga inkomster så långt tillbaka som 1968. Det innebär att vi kan studera hur olika personers inkomster har utvecklats över en väldigt lång tid.

I den här avhandlingen presenteras metoder för hur vi kan studera mönster i människors inkomstutveckling över tid med hjälp av funktionell dataanalys. Det innebär att vi betraktar inkomstkurvorna, specifikt kurvor av kumulativa livstidsinkomster, som kontinuerliga funktioner av tid. Detta förhållningssätt har stora konsekvenser för hur vi behandlar, analyserar och tolkar vårt data. Området funktionell dataanalys har utvecklats mycket de senaste decennierna, men det är fortfarande ovanligt att dessa metoder appliceras inom samhällsvetenskapen och på de frågor som tas upp i denna avhandling.

Det vi är intresserade av i denna avhandling är hur omständigheterna kring individers arbetsmarknadsinträde hänger ihop med deras framtida inkomstutveckling. Tjänar de som började jobba i en storstad i genomsnitt mer pengar än de som började jobba på landet? Hur utvecklas i så fall skillnaderna över loppet av individernas liv? Finns liknande skillnader mellan de som börjar jobba i ett nystartat respektive ett etablerat företag?

När vi studerar eventuella inkomstskillnader kan de inte utan vidare betraktas som orsakssamband. Vi kan till exempel inte påstå att inkomstskillnaderna faktiskt orsakas av var man börjar jobba. Det skulle nämligen kunna vara så att de personer som börjar jobba i olika lokala arbetsmarknader (eller olika typer av företag) också skiljer sig med avseende på andra aspekter som i sin tur påverkar deras inkomst, tex. utbildningsnivå eller föräldrarbakgrund. För att kunna dra slutsatser om eventuella orsakssamband från den här typen av data behöver vi använda oss av mera avancerade statistiska metoder, som hör till området kausal inferens. Ett stort bidrag i denna avhandling består i att vidareutveckla metoder för kausal inferens för att kunna användas på funktionella data.

Preface

Deras [ekonomernas] standardmodeller tänker sig nämligen aldrig att två personer övergivna på en öde ö kanske skulle börja prata med varandra, att de kanske skulle känna sig ensamma. Rädda. Behöva varandra. Efter att ha samtalat en stund skulle de inse att de bägge hade avskytt spenat när de var barn och hade farbröder som var alkoholiserade längre perioder. Efter att de hade resonerat en stund om detta skulle de antagligen dela på riset. Att vi människor kan reagera på det här sättet, har inte det ekonomisk betydelse?

- Katrine Marçal, Vem lagade Adam Smiths middag?

Words cannot express how grateful I am for all the support I have received during my time as a PhD-student. Anyway, here are several pages of words to express how grateful I am for all the support I have received during my time as a PhD-student:

It goes without saying that I could not have written this thesis without my supervisors, Xavier and Lina. Xavier, thank you for believing that I had what it takes to complete a PhD in statistics, even if I myself have questioned your judgement in this matter many, many times during the process[1]. Your ability to point out the most important thing to focus on, when there are way too many things I feel like I should be doing at the same time, has been invaluable to me finishing (somewhat) on time. And thank you, Lina, for always taking time to answer my spontaneous questions and help me with my code, even if the latter often entailed letting me explain what isn't working and finding the embarrassingly obvious mistake myself while doing so. Also, thank you for putting my feelings about my own understanding of the mathematical details of FDA into perspective. Thank you both for always giving me the appropriate level of support, from letting me figure things out (and make mistakes) on my own, to being there when I realised that I needed to ask for help. You have been amazing teachers, role-models and co-authors.

There is a figure of speech claiming that "it takes a village to raise a child"; that nurturing a small human life to maturity is a collaborative effort involving more people than just the child and their parents. I

[1] It's slowly starting to look like maybe you were right.

believe that "it takes a department to raise a PhD-student" would also be an appropriate saying. In many ways, getting to a point where I am about to defend the thesis you're now reading has felt like a collaborate effort involving far more people than just me and my supervisors. In many ways, I have also often felt like a child (or maybe the village idiot) when it comes to my knowledge about statistics[2]. But I have never been made to feel that way by anyone else than myself.

I have often seen academia in general be portrayed as a very competitive, almost hostile work environment. This may very well be the case in other countries, universities or departments. But my own experience has been the complete opposite. I am incredibly grateful to have been "raised" by coworkers who have been nothing but supportive to me and to each other. From celebrating each others' achievements with fika, to networks for support and collaboration for things like getting research funding or teaching courses, the focus has always been on collaboration rather than competition. From the day I started working at the department as a research assistant, I have always been treated like a full coworker, and I think that being taken seriously from the start has played a huge role to make me become a person who, to some extent, deserves to be taken seriously in an academic context.

So, to all of my coworkers at the department, thank you so much for modelling kindness and supportive behaviour in front of us academic children. Thank you for teaching me pretty much everything I know about statistics, both formally as my previous teachers and in informal contexts. Thank you for always trying to answer my questions, even the really weird ones. I honestly don't think I would have made it this far at a different department[3]. At the very least, I would not have enjoyed it nearly as much.

A special thanks also to my fellow PhD-students, past and present, for being an amazing peer-group and network of academic and social support. I have learned a lot from you, from practical aspects of what it means to be a PhD-student to diverse and important academic topics such as reinforcement learning, creating shiny apps in R, and very rude words in several different languages. I hope "growing up" together was just as much fun for you as it was for me.

I would also like to thank my co-authors, Olle Westerlund, Anna

[2] In my defence, almost everyone else has also been doing this for way longer than me.
[3] More on counterfactual outcomes in later parts of the thesis.

Baranowska-Rataj, Anna Brydsten, and Xijia Liu, for very interesting and rewarding collaborations. It has been a pleasure to work with you and I have learned a lot from you. Thank you to the academic resource center at the university library, and specifically Annika Bindler, for giving me the tools and inspiring me to become better at writing academic text. Also, I am incredibly grateful to Emmy and Sen for offering your professional help in making the thesis cover.

To my friends, thank you for your support and for giving me some much needed distractions from work, be it playing or officiating roller derby, walking dogs, playing DnD, or just spending time together. To Erik, for a frankly ridiculous amount of practical and emotional support, especially during the last part[4], thank you so much! You mean the world to me.

Umeå, December 2023
Kreske Ecker

[4]I.e, the last 5.5 years.

1 Introduction

Entry into the labour market is widely recognised as a crucial period of transition within life course analysis (see e.g., Russell and O'Connell, 2001; Scherer, 2001; McVicar and Anyadike-Danes, 2002; Brzinsky-Fay, 2007). The manner in which individuals (potentially) transition from education to employment, e.g. in terms of labour market attachment during youth and early adulthood, is considered to have a substantial impact on labour market outcomes later in the life course (Korpi et al., 2003; Brzinsky-Fay, 2007; Fauser, 2020). It is therefore important to follow up on the consequences of individuals' labour market entry conditions over extended periods of time, rather than to just consider immediate outcomes.

The availability of large scale, longitudinal administrative data in Sweden and other Nordic countries presents rare opportunities for researchers to monitor individuals' labour market outcomes over a large part of their (working) life course. At the same time, this type of data provides a rich set of measurements on potential early life determinants of labour market outcomes at the individual level.

In the articles included in this thesis, we apply and expand upon methods from functional data analysis (FDA) in order to analyse lifetime income trajectories. In general terms, FDA is concerning with data objects that are in the form of functions, e.g. curves, that vary over a domain such as time (see, e.g., Ramsay and Silverman, 2005; Wang et al., 2016; Aneiros et al., 2022, for an overview). Throughout this thesis, we analyse data that in practice is only measured at a finite number of discrete time points (e.g., annual observations of incomes from longitudinal registers). We can, however, regard those as discrete observations from an underlying continuous process, such as the continuous accumulation of income over the course of an individual's working life.

In many situations, it is of interest to draw conclusions about causal relationships between phenomena, such as the causal effects of initial labour market conditions on subsequent income trajectories. Doing so based on observational data, e.g. from administrative registers, requires more advanced methodology. In parts of this thesis, we develop methods to draw causal conclusions in the context of FDA.

In Paper I, we apply function-on-scalar linear regression and an interval wise testing procedure (Pini and Vantini, 2017; Abramowicz et al.,

2018) in order to study the associations between initial labour market size and lifetime income accumulation for one Swedish birth cohort. This approach reveals divergent patterns of income trajectories for different subgroups, but these associations cannot be interpreted as causal relationships. In Paper II, we present methods that allow for such causal interpretations by extending the potential outcome framework from causal inference to the context of FDA. This paper introduces a functional causal parameter of interest, the functional average treatment effect (FATE), as well as an estimator for the FATE. Inferences for the FATE are performed using simultaneous confidence bands (Liebl and Reimherr, 2023), and we present an application study on the causal effect of initial labour market size on cumulative income trajectories.

Paper III consists of an extended application of the methods introduced in Paper II to a different context. Here, we investigate the causal effect of initial firm age on subsequent earnings trajectories. The estimator of the FATE introduced in Paper II and further applied in Paper III is based on outcome regression (OR). We also show finite sample properties of this estimator in Paper II, conditional on some distributional assumptions about the underlying data. In Paper IV, we further derive the asymptotic distribution of this estimator under less restrictive assumptions, in the context of missing (functional) data. In addition, Paper IV also introduces a doubly robust (DR) estimator for the mean in the presence of missing data and presents its asymptotic properties.

1.1 Initial labour market conditions

Previous research within life course analysis highlights the importance of school-to-work transitions for outcomes such as labour market attachment and wage accumulation later in life. This body of research focuses primarily on micro-level conditions surrounding labour market entry, although they are understood to be embedded in larger institutional contexts. That is to say, the focus is on the individuals' labour market attachment in terms of temporary employment and unemployment, or participation in vocational training (Korpi et al., 2003; Brzinsky-Fay, 2007; Fauser, 2020).

A somewhat different perspective is to consider the surrounding macroeconomic conditions at the time of labour market entry, rather than labour market attachment at an individual level. Studies in this area find considerable associations between the timing of labour market entry and subsequent career outcomes (Altonji et al., 2016; Kahn, 2010; Kwon et al., 2010; Oreopoulos et al., 2012; Oyer, 2006; Raaum and Røed, 2006; Schwandt and von Wachter, 2019; Åslund and Rooth, 2007). The overall pattern emerging from these findings is that transitioning from school into the labour market during less favourable economic conditions, i.e. higher national unemployment rates, is associated with worse outcomes in terms of both subsequent wages and promotion rates. Most notably, these differences tend to persist for a substantial amount of time after labour market entry, even when macro-economic conditions have improved.

The explanations offered in this body of literature center around the notion that job matching efficiency is lower during economic downturns and new entrants into the labour market may be forced to accept employment in less desirable conditions. Consequently, they would spend the crucial first years of their careers in jobs that do not offer them opportunities to accumulate (the right kind of) human capital (Kahn, 2010; Oreopoulos et al., 2012). In a different research area, these two factors of job matching efficiency and accumulation of human capital have also been used to explain concurrent income differences between urban and rural labour markets. Here, the rationale is that urban labour markets not only provide an overall larger number of employment opportunities, but also offer workers better quality employer matches and more opportunities for skill formation and accumulation of human capital (Abel and Deitz, 2015; Wasmer and Zenou, 2002; Wheeler, 2001; De la Roca and Puga, 2017; Glaeser, 1999; Glaeser and Maré, 2001). Since very similar mechanisms have been proposed to explain the relationships between these different elements, this raises the question whether we also can find associations between geographic aspects of initial labour market conditions, namely the size of the initial labour market, and longer term outcomes in the form of subsequent income accumulation. This is examined empirically in Papers I and II in this thesis.

The above theories emphasise how the quality of (initial) jobs, in terms of opportunities for skill formation and accumulation of human capital, is important for workers' labour market trajectories. Another important question is how these aspects of job quality are related to characteristics of the place of employment. Here, firm age has been identified as one potentially relevant factor. Newly started firms generally tend to have access to fewer resources, not only in terms of financial capital but also knowledge, network connections and organisational routines (Coad et al., 2018). These factors are related to firms' economic performances, but they could also have consequences for the labour market outcomes of their workers. However, there is mixed empirical evidence as to whether employment in a new firm is associated with lower earnings by itself, or if earnings differences between new and older firms are actually due to characteristics of their workers (Brown and Medoff, 2003; Fackler et al., 2019, 2022; Nyström, 2021). This question is explored further in Paper III.

1.2 Register data

The empirical analyses in Papers I-III are conducted on register data. Register data is primarily collected for administrative purposes, such as taxation and the administration of health care and social services, rather than research. Despite that, the high degree of coverage, accuracy, and long historical availability make it a valuable data source for large-scale longitudinal and demographic analyses. In a global context, Sweden, along with other Nordic countries, stands out for maintaining extensive register data. This can be attributed to a long history of keeping demographic population registries as well as the comprehensive nature of the Nordic welfare state, which requires a large amount of administrative data to function efficiently. In addition, the system of unique personal identity numbers makes it easy to link records from different source registers. If such linkage is done for research purposes, it requires ethical approval.

More specifically, the data used in Papers I-III comes from the Umeå SIMSAM Lab infrastructure, which contains large amounts of longitudinal administrative data sourced from different Swedish registers and databases (Lindgren et al., 2016). This data comprises of individual-level socio-economic, demographic and health measures which are observed for the whole Swedish population over long time periods, as well as inter-generational links between family members.

2 Functional data analysis

The field of FDA collects methods for analysing phenomena that can be represented by functions of some continuous domain, most often of time (Ramsay and Silverman, 2005; Ferraty and Vieuw, 2006). The core units of analysis are thus entire functions, which, at least in theory, are infinite-dimensional. In reality however, it is neither possible to measure functions in a truly continuous manner, nor to store or analyse infinite-dimensional data. What we are able to observe in practice are a finite number of measurements based on some discretisation of the domain. Despite this discretisation, the paradigm of underlying continuous functions is central to setting FDA apart from other fields like multivariate statistics. That is, the focus is on the global behaviour of the curves, rather than their values at individual points, and the natural ordering of and dependency structure between observations contains valuable information in itself. The theoretical premise of infinite dimensional functional data objects has consequences for how statistical inference is approached in this field, which is discussed more extensively in Section 2.2.

There are two main ways to conceptualise functional data: as realisations of random elements that take values in a Hilbert space, or as sample paths of a stochastic process indexed on some closed interval of the real line (Hsing and Eubank, 2015). Throughout this text, we take the latter view and use the following notation for functional data:

$$y = \{y(t) : t \in \mathcal{T}\}, \mathcal{T} \subset \mathbb{R}.$$

Thus, we distinguish between the functional object, y, and $y(t)$ as the value this function takes at a specific point t. Without loss of generality, we can take the domain $\mathcal{T}$ to be $[0, 1]$.

2.1 Representation of functional data

Since functional data in practice only are observed at discrete points, it is often of interest to try to reconstruct the underlying continuous function from the discrete data. In many cases, the discrete observations also contain some level of measurement error or noise, which is assumed to not be present in the underlying function. The most common approach to reconstructing the underlying functions is to represent them by a linear combination of known basis functions ϕ_k, $k = 1, \ldots, K$; that is:

$$y(t) = \sum_{k=1}^{K} c_k \phi_k(t). \tag{1}$$

Popular choices of basis functions are Fourier bases for data that follows some periodical pattern (e.g. weather data) and B-spline bases for non-periodic data. Estimates of the coefficients c_k in (1) can be obtained by minimising a least squares criterion. The number of basis functions, K, determines how much the observed data points are smoothed rather than interpolated, with lower values of K corresponding to a larger degree of smoothing. Alternatively, the degree of smoothing can be controlled by explicitly introducing a smoothing parameter in the coefficient estimation (Ramsay and Silverman, 2005).

In the second paper, we apply a low degree of monotone smoothing to the raw data. Monotone smoothing restricts the derivative $Dy = dy/dt$ to be positive. This can be achieved by representing this derivative as the exponential of some unconstrained function W, yielding

$$Dy(t) = e^{W(t)}.$$

Here, W can be expressed as: $W(t) = \sum_{k=1}^{K} c_k \phi_k(t)$, using suitable basis function expansions. Unlike other smoothing approaches, monotone smoothing involves iteration, which makes it more computationally intensive (Ramsay and Silverman, 2005; Ramsay et al., 2009).

2.2 Inference for functional parameters

Many questions that arise in univariate statistical inference have clear analogues in the functional context. For instance, we may want to quantify the uncertainty in the estimate of a functional population parameter with the help of confidence regions, we may want to test whether the mean functions differ between two different populations, or whether a functional regression parameter is different from zero.

Questions such as the ones above can be interpreted as global statements concerning the behaviour of the functional parameter over its entire domain. That is to say, for a general unknown population parameter $\theta = \{\theta(t) : t \in \mathcal{T}\}$ we may want to test a hypothesis such as:

$$H_0 : \theta(t) = \theta_0(t) \ \forall t \in \mathcal{T}$$

against:

$$H_1 : \exists t \in \mathcal{T} : \theta(t) \neq \theta_0(t).$$

If this global null hypothesis is rejected, it is then of further interest to identify the point(s) or subsection(s) of the domain $\mathcal{T}$ for which the rejection occurred. This is referred to as local inference or domain selection.

One way of performing domain selection is to carry out pointwise inferences based on some discretisation of the domain, for example by constructing pointwise confidence bands at each point of this discretisation (e.g., Ramsay and Silverman, 2005; Reiss et al., 2010). The drawback of this approach is that it only provides a pointwise control of the type I error rate. Taken over the whole domain, the probability for a false positive would be much higher. This multiple comparison problem is a well known challenge in multivariate statistics, which has given rise to a number of different procedures aimed at adjusting for it. There are well-established procedures for controlling the family wise error rate (FWER), i.e. the probability of rejecting at least one true null hypothesis (Holm, 1979). Other procedures are aimed at controlling the false discovery rate (FDR), i.e. the rate of wrong rejections among all rejections of the null hypothesis (Benjamini and Hochberg, 1995; Olsen et al., 2021).

The fact that functional data are conceptualised as infinite dimensional in theory (and observed as high dimensional in practice) necessitates a different approach to the multiple comparison problem, since common adjustment techniques tend to loose power as the number of comparisons increases and cannot easily be extended to an infinite dimensional context from a conceptual point of view. In addition, characteristics inherent to functional data such as domain continuity, smoothness and local dependence structure, can provide valuable information to be used in domain selection (Pini and Vantini, 2017; Abramowicz et al., 2023).

One of the central challenges in FDA is thus to develop methods that are able provide domain selection as well as ensure a global control of the FWER (or FDR). Two main strategies for approaching this challenge are based on p-value adjustments and the construction of simultaneous confidence bands. The first strategy centers around performing a multitude of hypothesis tests and adjusting the resulting p-values to control the FWER. For example, the method presented in Vsevolozhskaya et al. (2014) utilises a closed testing procedure and is based on an a-priory partitioning of the functional domain. Pini and Vantini (2016) propose a procedure that is based on the coefficients from basis expansions of the functional data instead. However, these methods are sensitive to the choice of domain partition, or the choice of basis expansion, respectively. The interval-wise testing (IWT) procedure (Pini and Vantini, 2017; Abramowicz et al., 2018), which is described in detail in Section 2.2.1, further expands upon these principles. The aforementioned methods are non-parametric and thus do not make distributional assumptions about the data, but tend to be computationally demanding since they are based on permutations. More recently, Abramowicz et al. (2023) presented an attempt to unify different approaches of this kind within a common, more general framework.

Another strategy consists of constructing simultaneous confidence bands. Here, some methods are based on resampling techniques such as the parametric bootstrap (Degras, 2011; Wang et al., 2020). Other recent works build on geometric ideas (Choi and Reimherr, 2018) or results from random field theory (Telschow and Schwartzman, 2022; Liebl and Reimherr, 2023). The latter type tend to be less computationally demanding since they do not use resampling methods, but are instead based on distributional assumptions about the estimators and/or asymptotic results. The two specific methods applied in this thesis, IWT and one type of simultaneous confidence band, are described below.

2.2.1 Functional regression and interval wise testing procedure

Assume that we have a random sample $(y_i, \boldsymbol{x}_i),\ i = 1, \ldots, n$; where $y_i = \{y_i(t) : t \in [0, 1]\}$ are functional outcomes, and $\boldsymbol{x}_i = (1, x_{1i}, \ldots, x_{Li})^\top \in \mathbb{R}^L$ are vectors of scalar covariates. We assume $y_i \in C^0[0, 1]\ \forall i$. A function-on-scalar linear model is given by:

$$y_i(t) = \boldsymbol{x}_i^\top \boldsymbol{\beta}(t) + \varepsilon_i(t),$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_L(t))^\top$ are the functional regression coefficients and $\varepsilon_i(t)$ are i.i.d. random error functions with zero mean function $\mathbb{E}[\varepsilon(t)] = 0\ \forall t \in [0, 1]$, and finite total variance $\int_0^1 \mathbb{E}[\varepsilon(t)]^2\, \mathrm{d}t < \infty$.

The functional regression coefficients $\boldsymbol{\beta}(t)$ can be estimated using ordinary least squares (OLS) by minimising $\sum_{i=1}^n \int_0^1 \left(y_i(t) - \boldsymbol{x}_i^\top \boldsymbol{\beta}(t)\right)^2 \mathrm{d}t$

(Ramsay and Silverman, 2005). This yields the pointwise estimate:

$$\hat{\boldsymbol{\beta}}(t) = \underset{\beta_0(t),\ldots,\beta_L(t)}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i(t) - \boldsymbol{x}_i^\top \boldsymbol{\beta}(t) \Big)^2 \tag{2}$$

at each point t.

Given this model, it would be of interest to test whether a regression coefficient $\beta_l(t)$ is different from zero for some t and to identify the subsections of the domain for which that is the case. The principle of the IWT procedure (Pini and Vantini, 2017; Abramowicz et al., 2018) is to perform such domain selection by identifying the subintervals of the domain for which the null hypothesis is rejected.

For any closed interval $\mathcal{I} \subseteq [0,1]$ on the domain, the following hypotheses are tested:

$$H_{0,l}^{\mathcal{I}} : \beta_l(t) = 0, \ \forall t \in \mathcal{I}$$
$$H_{1,l}^{\mathcal{I}} : \beta_l(t) \neq 0, \ \text{for some } t \in \mathcal{I}.$$

This is done using functional permutation tests based on the Freedman and Lane permutation scheme (Freedman and Lane, 1983; Abramowicz et al., 2018), and with the test statistic $T_l^{\mathcal{I}} = \int_{\mathcal{I}} \Big(\hat{\beta}_l(t) \Big)^2 \mathrm{d}t$. This results in a p-value $p_l^{\mathcal{I}}$ corresponding to each interval $\mathcal{I}$. Then we define the adjusted p-value function as

$$\tilde{p}_l(t) = \sup_{\mathcal{I}: t \in \mathcal{I}} p_l^{\mathcal{I}}, \quad t \in [0,1],$$

i.e. as the supremum of the interval-wise p-values for all intervals containing t. We can then identify the sections of the domain for which $\beta_l(t)$ is significantly different from zero by thresholding the adjusted p-value function $\tilde{p}_l(t)$ at a given significance level.

The IWT procedure is consistent, and provides an interval-wise control of the type I error rate, that is to say,

$$\forall \mathcal{I} \subseteq [0,1] : H_{0,l}^{\mathcal{I}} \text{ is true} \implies \mathrm{P}[\forall t \in \mathcal{I}, \tilde{p}_l(t) \leq \alpha] \leq \alpha,$$

$\forall \alpha \in [0,1]$ (Pini and Vantini, 2017; Abramowicz et al., 2018).

2.2.2 Simultaneous confidence bands

The other method for drawing inferences regarding functional parameters that is applied in this thesis involves the construction of simultaneous confidence bands. This is used in Papers II-IV. Specifically, we apply the "Fast and fair simultaneous confidence bands" presented in Liebl and Reimherr (2020, 2023). Their works draw upon concepts from random field theory, namely the expected Euler characteristic and generalisations of the Kac-Rice formulas, to construct confidence bands with adaptive, non-constant widths. These bands are fair in the sense that false positive rates are balanced across arbitrary subsections of the domain. This allows for local interpretations within these subsections, while maintaining a global control of coverage probabilities. The construction of these bands is generally less computationally intensive compared to methods that are based on permutations and resampling (Liebl and Reimherr, 2023).

Let $\theta = \{\theta(t) : t \in [0, 1]\}$ be an unknown functional parameter with $\theta \in C^1[0, 1]$ almost surely. Let $\hat{\theta} = \{\hat{\theta}(t) : t \in [0, 1]\}$ be an estimator satisfying

$$\sqrt{r_n}\left(\hat{\theta}(t) - \theta(t)\right) \xrightarrow{d} \mathcal{N}\left(0, C_\theta(t,t)\right) \ \forall t \in [0,1]$$

as $n \to \infty$, where $\sqrt{r_n}$ is the convergence rate of the estimator and $C_\theta(s,t) = \mathrm{Cov}\left(\hat{\theta}(s), \hat{\theta}(t)\right)$, $s, t \in [0,1]$, its covariance function.

A valid $100 \times (1-\alpha)\%$ simultaneous confidence band for $\theta(t)$ is given by

$$\mathrm{SCB}(t) = \hat{\theta}(t) \pm u^*_{\alpha/2}(t)\sqrt{C_\theta(t,t)/r_n}.$$

The critical value function $u^*_{\alpha/2}(t)$ is not constant and thus able to adapt to the local dependence structure of the data. In general, there is more than one possible critical value function guaranteeing a global control of false positive rates, $\mathrm{P}\left(\exists t \in [0,1] : \theta(t) \notin [\mathrm{SCB}_l(t), \mathrm{SCB}_u(t)]\right) \leq \alpha$, so that $u^*_{\alpha/2}(t)$ can be chosen among them with respect to the fairness constraint of balancing the false positive rates over different domain partitions (Liebl and Reimherr, 2023).

Figure 1 illustrates the adaptive critical value function and the simultaneous confidence bands. The data is based on the Monte Carlo simulations in Paper IV, where the goal is to estimate the mean function in the presence of missing functional outcomes. Some of the simulated outcome functions and their mean, observed at 50 time points on $[0, 1]$, are shown in the top row. The middle row shows the adaptive critical value function $u^*_{0.025}(t)$, based on a partition of the domain into six equidistant subsections. The bottom graph compares the width of the

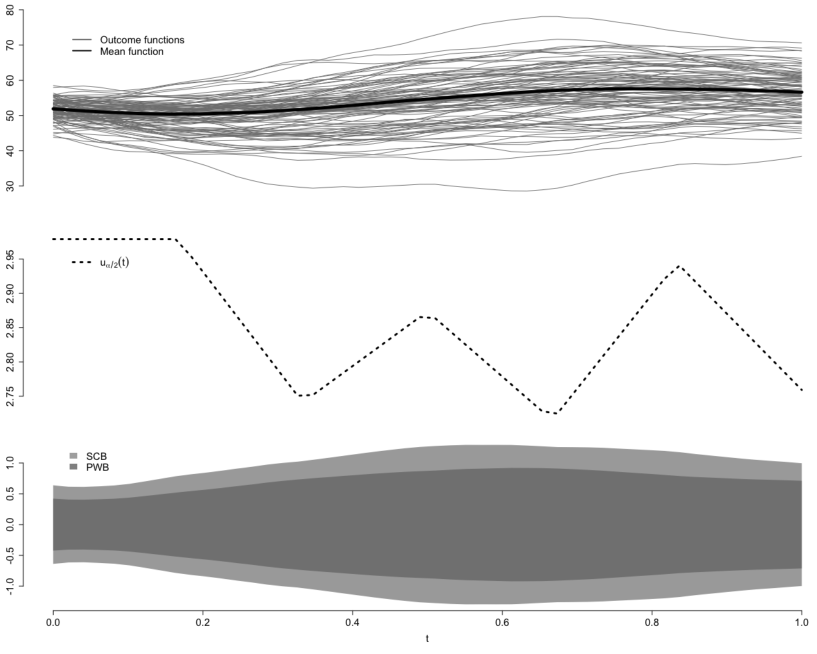

Figure 1: Top row: Simulated outcome functions (grey) and mean function (black) from the simulation study in Paper 4. Middle row: Adaptive critical value function $u^*_{\alpha/2}(t)$. Bottom row: Centered simultaneous confidence bands (SCB, light grey) and pointwise confidence bands (PWB, dark grey) for the mean function.

resulting 95% simultaneous confidence bands for the mean function with their pointwise equivalents. The latter are constructed using a constant critical value function $\tilde{u}_{0.025}(t) = 1.96,\ \forall t \in [0, 1]$, and do not account for the multiple comparison problem. The bands have been centered to ease the comparison of their width.

3 Causal inference

Section 1.1 discussed the associations between initial labour market conditions and subsequent outcomes such as earnings. The phrase 'association' is used intentionally to emphasise that these statements only concern the correlation between two phenomena, which does not necessarily imply a causal relationship. In many situations however, we are interested in drawing conclusions about causal relationships as well. Namely, we want to study the effect that a certain treatment, or exposure to certain conditions (such as entering the labour market in a certain geographic area or type of firm) has on an outcome (such as subsequent income accumulation).

Randomised experiments are considered the gold standard for identifying causal effects of treatments. In the simplest form, participants are randomly divided into two groups and are either assigned the treatment or belong to the control group. Since treatment assignment is randomised and thus not affected by characteristics of the individuals in question, the average differences in the outcome between the two groups can be used as an estimate of the treatment's effect.

There are many situations in which the use of randomised experiments would be neither ethical nor practically feasible. For the topics considered in this thesis for example, it would not be reasonable to randomly assign individuals to start working in a certain location, or to seek employment in a certain type of firm. The observational nature of the data in question makes the identification of causal effects more difficult.

When the treatment assignment is not randomised, there is a risk that the causal effect is confounded by differences in the composition of the treatment groups. The individuals who receive a treatment (e.g. enter the labour market in an urban area) could differ systematically from the individuals in the control group with respect to characteristics (such as education level) that in turn are also associated with the outcome.

One way to formalise the study of causal effects is to use the potential outcome framework. This framework was first introduced in an experimental setting by Neyman in 1925 (Splawa-Neyman et al., 1990), but gained additional traction when applied to the context of observational data by Rubin (1974). In this thesis, we study the effects of some (scalar) treatment on outcomes that are functional in nature. A central part of this work is to extend the potential outcome framework to the context of functional data.

3.1 Extending the potential outcome framework to functional data

We are interested in the causal effect that a treatment has on a functional outcome (i.e., a function of time). The treatment is assumed to be binary and occur at one specific time point. In extension, this means that the causal effect also is a function of time. In order to discuss this more formally, we introduce the following notation.

For each individual i in the study, $i = 1, \dots, n$, we have a (scalar) binary treatment indicator:

$$z_i = \begin{cases} 1, & \text{if treated,} \\ 0, & \text{if not treated.} \end{cases}$$

We then define two potential outcome functions for each unit, which correspond to these two treatment levels. Let $y_{1i} = \{y_{1i}(t) : t \in [0, 1]\}$ be the outcome function that individual i would have displayed if they had received the treatment, and $y_{0i} = \{y_{0i}(t) : t \in [0, 1]\}$ be the outcome function that individual i would have displayed if they had not received the treatment. We assume that the treatment assignment of one individual does not affect the potential outcomes of other individuals in the study, and that each unit receiving the treatment receives the same version of it. In practice, only one of the two potential outcome functions is realised for each person, and we denote the observed outcome for individual i by

$$y_i = y_{1i}\, z_i + y_{0i}\, (1 - z_i).$$

In addition, each individual also has a vector of covariates, $\boldsymbol{x}_i = (1, x_{1i}, \dots, x_{Li})^\top$ associated with them, which are measured before the treatment assignment.

The parameter of interest is the functional average treatment effect (FATE), defined as:

$$\theta := \{\theta(t) = \mathbb{E}\left(y_{1i}(t) - y_{0i}(t)\right) : t \in [0, 1]\}\,. \tag{3}$$

Notably, the FATE is defined in (3) in terms of both potential outcomes for each unit, only half of which are actually observed. In order to be able to identify and estimate the FATE, we therefore need to impose additional assumptions on the observational data (e.g., Wooldridge, 2010).

3.1.1 Identification Assumptions

We make the following assumptions:
Assumption 1 (Ignorability in mean)

$$
\begin{aligned}
\mathbb{E}\left(y_{0i}(t) \mid z_i, \boldsymbol{x}_i\right) &= \mathbb{E}\left(y_{0i}(t) \mid \boldsymbol{x}_i\right) \qquad \forall t \in [0,1], \\
\mathbb{E}\left(y_{1i}(t) \mid z_i, \boldsymbol{x}_i\right) &= \mathbb{E}\left(y_{1i}(t) \mid \boldsymbol{x}_i\right) \qquad \forall t \in [0,1].
\end{aligned}
$$

This first assumption holds if there are no unobserved confounders related to both the potential outcomes and the treatment assignment. To the extent that such confounders exist, we assume that they are measured by the observed covariate vector $\boldsymbol{x}_i$. This is a slightly weaker version of the ignorability assumption, which states that potential outcomes jointly are independent of the treatment assignment given the observed covariates, $(y_{0i}(t), y_{1i}(t)) \perp\!\!\!\perp z_i \mid \boldsymbol{x}_i$. For many purposes, including what is covered in this thesis, Assumption 1 is sufficient.

The assumption of mean ignorability cannot be tested empirically and the potential existence of unobserved confounders is a common challenge when trying to draw causal conclusions from observational data. One way of addressing this challenge is to test the robustness of the results to violations of this assumption. This can for example be done using uncertainty intervals (Vansteelandt et al., 2006; Genbäck and de Luna, 2019). In addition, domain knowledge can be used to judge the plausibility of this assumption in a specific context.

Assumption 2 (Overlap)

$$
\begin{aligned}
\mathrm{P}(z_i = 0 \mid \boldsymbol{x}_i) > 0, \quad \forall \boldsymbol{x}_i, \\
\mathrm{P}(z_i = 1 \mid \boldsymbol{x}_i) > 0, \quad \forall \boldsymbol{x}_i.
\end{aligned}
$$

The overlap assumption implies that it is possible for all individuals to receive either treatment regime, regardless of their values on the observed covariates. This is also sometimes referred to as common support. The validity of the overlap assumption can be checked by comparing the empirical distributions of the treatment probability between the treatment groups. In the case of violations, it is often possible to redefine the populations of interest so that the overlap assumption is fulfilled.

Given these assumptions, the functional causal parameter θ can be identified as follows (see Paper II):

$$\begin{aligned}\theta(t) &= \mathbb{E}(y_{1i}(t)) - \mathbb{E}(y_{0i}(t))\\ &= \mathbb{E}\big[\mathbb{E}(y_{1i}(t) \mid z_i = 1, \boldsymbol{x}_i) - \mathbb{E}(y_{0i}(t) \mid z_i = 0, \boldsymbol{x}_i)\big]\\ &= \mathbb{E}\big[\mathbb{E}(y_i(t) \mid z_i = 1, \boldsymbol{x}_i) - \mathbb{E}(y_i(t) \mid z_i = 0, \boldsymbol{x}_i)\big], \qquad \forall t \in [0,1].\end{aligned}$$

3.2 Estimators of the FATE

In Papers II - IV, we develop and apply two estimators for the FATE. We present an outcome regression (OR) estimator and a double robust (DR) estimator, which are based on models for the relationship between the potential outcomes and the covariates, and between the treatment assignment and the covariates, respectively. For the potential outcomes, we assume a function-on-scalar linear model (Ramsay and Silverman, 2005):

$$\begin{aligned}y_{0i}(t) &= \boldsymbol{x}_i^\top \ \boldsymbol{\beta}_0(t) + \varepsilon_{0i}(t),\\ y_{1i}(t) &= \boldsymbol{x}_i^\top \ \boldsymbol{\beta}_1(t) + \varepsilon_{1i}(t),\end{aligned} \tag{4}$$

where $\boldsymbol{\beta}_z(t)$, $z = 0, 1$, are $L+1$-dimensional functional parameters, and the random errors $\varepsilon_{zi} = \{\varepsilon_{zi}(t) : t \in [0,1]\}$ are assumed to be independent stochastic processes with zero mean functions and finite covariance functions $\sigma_z(s,t) = \operatorname{Cov}\big(\varepsilon_{zi}(s), \varepsilon_{zi}(t)\big)$, $s, t \in [0,1]$, with $\sigma_z(t,t) > 0$.

For the treatment assignment, we model the propensity score $\pi(\boldsymbol{x}_i) = \mathrm{P}(z_i = 1 \mid \boldsymbol{x}_i)$, for example using a logistic regression:

$$\pi(\boldsymbol{x}_i) = \left(1 + e^{\boldsymbol{x}_i^\top \boldsymbol{\gamma}}\right)^{-1}, \tag{5}$$

where $\boldsymbol{\gamma}$ is a $(L+1) \times 1$ vector of coefficients.

Based on the model in (4), the FATE can be estimated using OR imputation as:

$$\hat{\theta}_{OR}(t) = \frac{1}{n} \sum_{i=1}^{n} \Big(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_1(t) - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_0(t)\Big), \tag{6}$$

where $\hat{\boldsymbol{\beta}}_1(t)$ and $\hat{\boldsymbol{\beta}}_0(t)$ can be obtained by pointwise regression of the observed outcome on the covariates within the subsets of treated and controls, respectively, e.g. as in (2).

Based on the models in (4) and (5), a DR estimator of the FATE can be specified as follows:

$$\hat{\theta}_{DR}(t) = \hat{\theta}_{OR}(t) + \frac{1}{n}\sum_{i=1}^{n} \frac{z_i\, e_{1i}(t)}{\hat{\pi}(\boldsymbol{x}_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1-z_i)\, e_{0i}(t)}{1-\hat{\pi}(\boldsymbol{x}_i)}, \qquad (7)$$

where $e_{zi}(t) = y_i(t) - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_z(t)$ are the residuals from the two outcome regression models and $\hat{\pi}(\boldsymbol{x}_i)$ are estimates of the propensity score obtained by e.g. maximum likelihood estimation.

Under certain conditions, the OR estimator in (6) has a Gaussian process as its finite sample distribution, as is shown in Paper II. In Paper IV, we show that both this OR estimator and the DR estimator in (7) asymptotically are Gaussian processes even under more general conditions. The double robustness property of the DR estimator means that it is consistent and asymptotically Gaussian even if one of the two models in (4) and (5) is specified incorrectly.

In order to quantify the uncertainty in these estimators and to perform inferences about the FATE, we can for example use the simultaneous confidence bands described in Section 2.2.2, which is done in Papers II-IV. However, inference based on the estimators presented above is not limited to these specific methods, but can be performed using other techniques as well. Since the estimators are asymptotically Gaussian, this allows for inferential methods that rely on distributional assumptions about the estimators.

3.2.1 Missing data

The previous sections extended the potential outcome framework of causal inference to the context of FDA, and introduced the FATE as well as estimators of this parameter. The central concepts of the potential outcome framework are closely related to the topic of estimating outcomes in the presence of missing data, specifically data that is missing at random (Ding and Li, 2018). Since only one of the two potential outcomes is realised for each individual, we can view the other, counterfactual outcome as being missing for that person. This means that analogues of the estimators of the FATE in (6) and (7) can be used in the context of missing data, which is done in Paper IV. Specifically, we present estimators for the mean of functional outcomes when some of these outcomes are missing at random. That is to say, we assume that the missingness depends on the value of some covariates, which are scalars and always observed, but not on the value of the functional outcomes themselves.

4 Summary of papers

4.1 Paper I

Previous research has found associations between macro-economic conditions at labour market entry and subsequent wage- and career trajectories. Other studies describe associations between concurrent labour market size and earnings, referred to as urban wage premiums. In the first paper, we aim to connect these two research areas and examine the relationship between initial labour market size and subsequent trajectories of cumulative income. We model the cumulative incomes of the 1954 Swedish birth cohort as functions of time, and examine how these income trajectories are linked to initial labour market size using function-on-scalar regression models. We also control for background factors such as gender, education level and variables related to socio-economic background. Inference about the functional regression coefficients is performed using the IWT procedure. By using methods from functional data analysis, we can examine how the relationships between income functions and the explanatory variables change dynamically over time.

The results show divergent patterns in lifetime income trajectories among different groups. Men generally accumulate more income over their lifetime if they start working in urban areas, and this difference grows as they age. However, this pattern does not hold for men with only primary education, who instead tend to earn less if they begin their careers in big cities. This gap between men with different education levels widens over time. On the other hand, women of all educational backgrounds benefit from starting their careers in large urban areas, with the largest income advantage seen for those with only primary education. For women, the differences between education levels diminish over time.

4.2 Paper II

In the second paper, we introduce methods for causal inference in the context of FDA. The motivating practical example concerns estimating the effect of initial labour market size (a binary treatment) on subsequent cumulative income trajectories (a functional outcome). We define the parameter of interest, the FATE, as well as develop an estimator based on OR imputation, by extending the potential outcome framework to a functional setting. Furthermore, we establish the finite sample distribution of this estimator under certain conditions and show how simultaneous confidence bands can be used to obtain valid inferences about the FATE. The properties of our estimator and the accompanying confidence bands are illustrated in a simulation study.

A practical application study is based on the motivating example. Here we find that living in an urban area at the age of 20 has a positive causal effect on cumulative lifetime incomes overall. However, we find different patterns in the effect when performing separate analyses for men and women. For men, there is an initial negative effect, which becomes positive over time. For women, there is a stronger effect, which is positive over the entire study period.

4.3 Paper III

The third paper investigates how the age of their initial place of employment impacts the long-term earnings of labour market entrants in Sweden. Previous research has come to mixed conclusions in this matter. We ask if there is a causal effect of firm age on earnings trajectories even when controlling for confounding variables, or if earnings differences between new and established firms can be explained by differences in workers' background characteristics. Another central question is whether such an effect persists over time. Utilising Swedish longitudinal register data, the study follows two cohorts who entered the job market in 1995 and 1999 over an extended period of time (17-21 years). We employ the methods introduced in the previous paper, which allow us to examine how the functional causal effect of firm age on earnings accumulation changes over time.

The findings indicate a negative impact on the earnings accumulation of individuals who began their careers in new firms in 1999, compared to those who started in more established firms. This effect is present even when a large number of relevant background characteristics are controlled for, and persists over time. It is strongest a few years after labour market entry. However, for the 1995 labour market entry cohort, this effect on earnings accumulation is not significant. While we do find raw differences in earnings between the two treatment groups in this sub-population, these can be explained by differences in workforce composition between new and older firms.

4.4 Paper IV

Missing data is a well-known problem, which can arise in many different situations. We say that outcome data is missing at random (MAR) when the missingness depends on the values of some fully observed covariates, but not on the outcomes themselves. In the fourth paper, we present methods to deal with MAR functional outcomes. We present semi-parametric OR and DR estimators for the mean of functional outcomes in this scenario.

Our theoretical results show that both estimators asymptotically are Gaussian processes and establish the double robustness property of the DR estimator in the functional setting. We also demonstrate the use of simultaneous confidence bands for drawing inferences about the mean function. The theoretical properties of the two estimators and the coverages of their corresponding simultaneous confidence bands are confirmed in Monte Carlo simulations.

5 Final remarks and further research

This thesis makes two important contributions. First, we showcase how methods from FDA can be used to model labour market trajectories from a life course perspective. Practical applications of FDA methods are often based on subject matter from the natural and medical, rather than social sciences (see Ullah and Finch, 2013; Aneiros et al., 2019a,b, for overviews), although there are some recent applications using financial or econometric data (e.g., Wang et al., 2008; Kneip and Liebl, 2020; Zhang and Li, 2022). As far as we know, Papers I-III are the first to apply FDA methods to individual earnings trajectories and to study divergences in these functional trajectories based on mechanisms of social stratification and initial labour market conditions from a life course perspective.

The second contribution consists of developing methods to draw causal inferences for functional outcomes, which is done in Papers II and IV. To our knowledge, there is only one earlier attempt at studying causal inference for a functional outcome. Belloni et al. (2017) introduce double robust estimators and their limiting distributions, albeit within a much more general context of high dimensional causal inference. Their approach differs from ours in that they do not present expressions of the covariance structure of these estimators and use inferential techniques that rely on bootstrap.

The OR and DR estimators presented in this thesis involve nuisance models for the relationships between the covariates and the outcome functions, as well as the relationships between the covariates and the treatment assignment or missingness mechanism. We have estimated the parameters in these nuisance models using function-on-scalar linear regression and logistic regression, respectively. These types of models are comparatively simple, although well established and frequently used in the scalar case as well. A natural next step for research in this area would thus be to extend more advanced approaches to nuisance model estimation to the functional framework, for example targeted maximum likelihood estimation (van der Laan and Rose, 2018) or methods for high dimensional nuisance models (see Moosavi et al., 2023, for an overview). It would also be of interest to expand the framework for functional causal inference derived in this thesis to allow for time-varying treatment assignments and/or functional covariates, which in turn would involve more complicated nuisance models.

In the practical applications of our methods, we have studied the associations between initial labour market conditions and earnings trajectories during parts of, rather than the entirety of individuals' life courses. The length for which we can follow up on individuals' earnings after labour market entry is limited by data availability. Once longer follow up periods of data are available in the future, a relevant topic to consider would be to what extent initial divergences in income trajectories translate into inequality in later stages of the life course, e.g., into inequalities in economic outcomes in retirement.

References

Abel, J. R. and Deitz, R. (2015). Agglomeration and job matching among college graduates. *Regional Science and Urban Economics*, 51(4):14–20.

Abramowicz, K., Häger, C. K., Pini, A., Schelin, L., Sjöstedt de Luna, S., and Vantini, S. (2018). Nonparametric inference for functional-on scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scandinavian Journal of Statistics*, 45(4):1036–1061.

Abramowicz, K., Pini, A., Schelin, L., Sjöstedt de Luna, S., Stamm, A., and Vantini, S. (2023). Domain selection and familywise error rate for functional data: A unified framework. *Biometrics*, 79(2):1119–1132.

Altonji, J. G., Kahn, L. B., and Speer, J. D. (2016). Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success. *Journal of Labor Economics*, 34(S1):S361–S401.

Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019a). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9.

Aneiros, G., Cao, R., and Vieu, P. (2019b). Editorial on the special issue on functional data analysis and related topics. *Computational Statistics*, 34:447–450.

Aneiros, G., Horová, I., Hušková, M., and Vieu, P. (2022). On functional data analysis and related topics. *Journal of Multivariate Analysis*, 189:104861.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233 – 298.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1):289–300.

Brown, C. and Medoff, J. L. (2003). Firm age and wages. *Journal of Labor Economics*, 21(3):677–697.

Brzinsky-Fay, C. (2007). Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe. *European Sociological Review*, 23(4):409–422.

Choi, H. and Reimherr, M. (2018). A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1):239–260.

Coad, A., Holm, J. R., Krafft, J., and Quatraro, F. (2018). Firm age and performance. *Journal of Evolutionary Economics*, 28:1–11.

De la Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *Review of Economic Studies*, 84(1):106–142.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765.

Ding, P. and Li, F. (2018). Causal inference: A missing data perspective. *Statistical Science*, 33(2):214–237.

Fackler, D., Fuchs, M., Hölscher, L., and Schnabel, C. (2019). Do start-ups provide employment opportunities for disadvantaged workers? *ILR Review*, 72(5):1123–1148.

Fackler, D., Hölscher, L., Schnabel, C., and Weyh, A. (2022). Does working at a start-up pay off? *Small Business Economics*, 58:2211–2233.

Fauser, S. (2020). Career trajectories and cumulative wages: The case of temporary employment. *Research in Social Stratification and Mobility*, 69:100529.

Ferraty, F. and Vieuw, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. New York: Springer, 1st edition.

Freedman, D. and Lane, D. (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business and Economic Statistics*, 1(4):292–298.

Genbäck, M. and de Luna, X. (2019). Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. *Biometrics*, 75(2):506–515.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*, 46(2):254–277.

Glaeser, E. L. and Maré, D. (2001). Cities and Skills. *Journal of Labour Economics*, 19(2):316–342.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons, 1st edition.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316.

Kneip, A. and Liebl, D. (2020). On the optimal reconstruction of partially observed functional data. *The Annals of Statistics*, 48(3):1692–1717.

Korpi, T., De Graaf, P., Hendrickx, J., and Layte, R. (2003). Vocational training and career employment precariousness in Great Britain, the Netherlands and Sweden. *Acta Sociologica*, 46(1):17–30.

Kwon, I., Meyersson Milgrom, E., and Hwang, S. (2010). Cohort Effects in Promotions and Wages Evidence from Sweden and the United States. *Journal of Human Resources*, 45(3).

Liebl, D. and Reimherr, M. (2020). Simultaneous Inference for Function-valued Parameters: a Fast and Fair Approach. In Aneiros, G., Horová, I., Hušková, M., and Vieu, F., editors, *Functional and High-Dimensional Statistics and Related Fields*. Cham: Springer Nature Switzerland.

Liebl, D. and Reimherr, M. (2023). Fast and fair simultaneous confidence bands for functional parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):842–868.

Lindgren, U., Nilsson, K., de Luna, X., and Ivarsson, A. (2016). Data Resource Profile: Swedish Microdata Research from Childhood into Lifelong Health and Welfare (Umeå SIMSAM Lab). *International Journal of Epidemiology*, 45(4):1075–1075g.

McVicar, D. and Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 165(2):317–334.

Moosavi, N., Häggström, J., and de Luna, X. (2023). The costs and benefits of uniformly valid causal inference with high-dimensional nuisance parameters. *Statistical Science*, 38(1):1–12.

Nyström, K. (2021). Working for an entrepreneur: heaven or hell? *Small Business Economics*, 56:919–931.

Olsen, N. L., Pini, A., and Vantini, S. (2021). False discovery rate for functional data. *Test*, 30(3):784–809.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The short and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29.

Oyer, P. (2006). Initial Labor Market Conditions and Long-Term Outcomes for Economists. *Journal of Economic Perspectives*, 20(3):143–160.

Pini, A. and Vantini, S. (2016). The interval testing procedure: a general framework for inference in functional data analysis. *Biometrics*, 72(3):835–845.

Pini, A. and Vantini, S. (2017). Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, 29(0):835–845.

Raaum, O. and Røed, K. (2006). Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects? *The Review of Economics and Statistics*, 88(2):193–210.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. New York: Springer.

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

Reiss, P. T., Huang, L., and Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6(1):1–26.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Russell, H. and O'Connell, P. J. (2001). Getting a job in europe: The transition from unemployment to work among young people in nine european countries. *Work, Employment and Society*, 15(1):1–24.

Scherer, S. (2001). Early Career Patterns: A Comparison of Great Britain and West Germany. *European Sociological Review*, 17(2):119–144.

Schwandt, H. and von Wachter, T. (2019). Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets. *Journal of Labor Economics*, 37(S1):S161–S198.

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on principles. Section 9. *Statistical Science*, 5(4):465–472.

Telschow, F. and Schwartzman, A. (2022). Simultaneous confidence bands for functional data using the Gaussian Kinematic Formula. *Journal of Statistical Planning and Inference*, 216(01):70–94.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13:1–12.

van der Laan, M. J. and Rose, S. (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Cham: Springer.

Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979.

Vsevolozhskaya, O., Greenwood, M., and Holodov, D. (2014). Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *The Annals of Applied Statistics*, 8(2):905–925.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional Data Analysis. *The Annual Review of Statistics and Its Application*, 3(1):257–295.

Wang, S., Jank, W., and Shmueli, G. (2008). Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business & Economic Statistics*, 26(2):144–160.

Wang, Y., Wang, G., Wang, L., and Ogden, R. T. (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics*, 76(2):427–437.

Wasmer, E. and Zenou, Y. (2002). Does city structure affect job search and welfare? *Journal of Urban Economics*, 51(3):515–541.

Wheeler, C. (2001). Search, Sorting, and Urban Agglomeration. *Journal of Labor Economics*, 19(4):879–899.

Wooldridge, J. F. (2010). *Econometric Analysis of Cross Section and Panel Data, second edition.* Cambridge: Massachusetts Institute of Technology.

Zhang, H. and Li, Y. (2022). Unified principal component analysis for sparse and dense functional data under spatial dependency. *Journal of Business & Economic Statistics*, 40(4):1523–1537.

Åslund, O. and Rooth, D. (2007). Do when and where matter? Initial labour market conditions and immigrant earnings. *The Economic Journal*, 117(3):422–448.