Journal of the Royal Statistical Society Series C: Applied Statistics, 2024, **73**, 221–240
https://doi.org/10.1093/jrsssc/qlad092
Advance access publication 26 October 2023
Original Article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Causal inference with a functional outcome

Kreske Ecker, Xavier de Luna and Lina Schelin

Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden

Address for correspondence: Kreske Ecker, Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Biblioteksgränd 6, 90187 Umeå, Sweden. Email: kreske.ecker@umu.se

Abstract

This article presents methods to study the causal effect of a binary treatment on a functional outcome with observational data. We define a Functional Average Treatment Effect (FATE) and develop an outcome regression estimator. We show how to obtain valid inference on the FATE using simultaneous confidence bands, which cover the FATE with a given probability over the entire domain. Simulation experiments illustrate how the simultaneous confidence bands take the multiple comparison problem into account. Finally, we use the methods to infer the effect of early adult location on subsequent income development for one Swedish birth cohort.

Keywords: early adult location, functional average treatment effect, lifetime income trajectory, simultaneous confidence bands

1 Introduction

The impact of economic conditions at labour market entry on subsequent wage development and other labour market outcomes has been the focus of a range of recent studies; see e.g. Altonji et al. (2016), Åslund and Rooth (2007), Kwon et al. (2010), Oreopoulos et al. (2012), Raaum and Røed (2006), and Schwandt and von Wachter (2019). In this article, we contribute by utilising large-scale register data providing much longer follow-up periods, and by developing methods that allow us to draw valid inference in this context, i.e. taking into account an inherent multiple comparison problem regarding the causal effect of initial labour market conditions on income observed through whole working lives.

More precisely, we investigate the causal effect of early adult residence location on subsequent cumulative lifetime incomes for the 1954 Swedish birth cohort, based on population-wide register data. The data available allow us to follow income development through most of the cohort's working life: from adolescence until the age of 63. We define the binary treatment of interest as living in either an urban or a rural area of Sweden at the age of 20 (in 1974). The outcome, logarithmised cumulative income (LCI), is observed annually. Figure 1 shows the mean differences in LCI between the two treatment groups (urban–rural location at age 20), computed separately for men and women, over the period 1975–2017, together with 95% pointwise confidence bands. This comparison is naive in two aspects: first, it has no causal interpretation since confounding variables are not controlled for; and second, the pointwise confidence bands are not meaningful because of the multiple comparison problem, i.e. the displayed bands yield only pointwise control of the coverage probability, whereas the actual nominal coverage over the entire domain of the income functions is considerably lower.

The observed incomes are assumed to arise from an underlying continuous income accumulation process, from which we take discrete observations. This motivates the use of methods from

Received: January 26, 2023. **Revised:** September 14, 2023. **Accepted:** September 15, 2023
© The Royal Statistical Society 2023.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

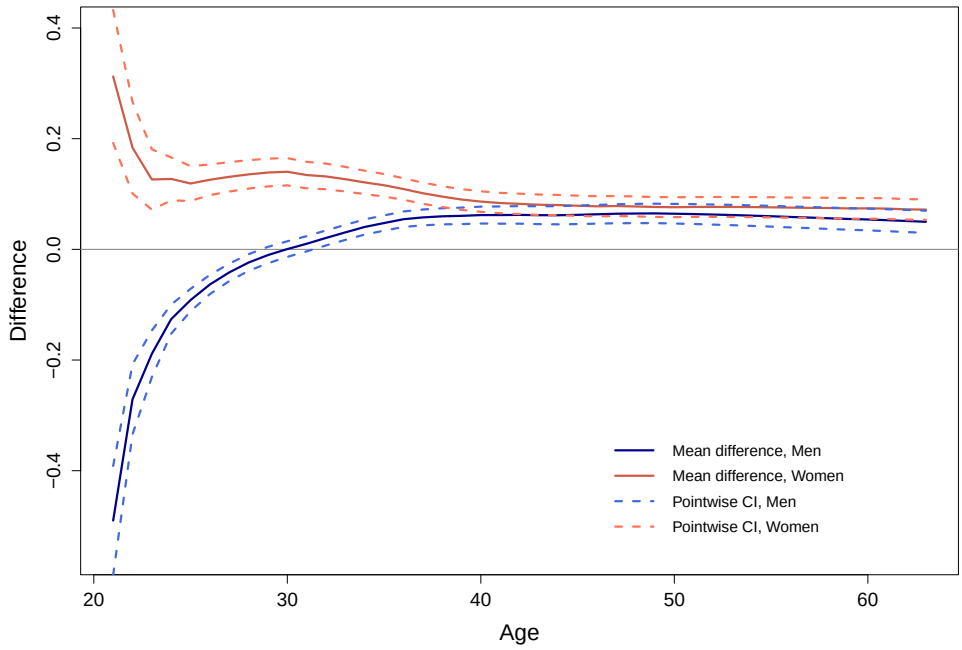

Figure 1. Mean differences in LCI (urban versus rural location at age 20), together with 95% pointwise confidence bands, for the 1954 birth cohort in Sweden. By gender.

the field of functional data analysis (FDA) (see, e.g. Ramsay & Silverman, 2005; Wang et al., 2016, for an overview). While FDA methods are not the only way to analyse such data, they have a major advantage in that they allow to address the multiple comparison problem in an elegant and practical fashion as described below. To deal with a non-randomised treatment assignment (here early adult residence location), we use the potential outcome framework (Holland, 1986; Rosenbaum & Rubin, 1983) to define a causal parameter, the Functional Average Treatment Effect (FATE), which takes the functional nature of the outcome into account. Up to our knowledge, there is only one earlier attempt (Belloni et al., 2017) to formalise and study causal inference in the context of a functional outcome. Our methodological contribution differs in that we provide exact finite sample inference, while Belloni et al. (2017) provide justification for asymptotic bootstrap inference. Other related works include Chib & Jacobi (2007) and Jacobi et al. (2016) in which the terminology 'panel outcomes' is used. In this literature, the multiple comparison problem is not dealt with, although it is also less severe since typically only a few time points of follow up are considered. Note that complex causal effects can also be defined when scalar potential outcomes are compared, e.g. by contrasting their distribution functions (Lin et al., 2021).

We propose and study an estimator of the FATE based on outcome regression (Tan, 2007). Quantifying the uncertainty in this estimation presents a challenge since existing inferential techniques developed for univariate outcomes cannot satisfactorily address the multiple comparison problem induced by the functional nature of the causal parameter in this setting. Since the functional parameter is infinite-dimensional, standard techniques to control for multiple testing (e.g. Holm, 1979) would be too conservative. Instead, techniques within the field of FDA provide a variety of solutions to the multiple comparisons problem, including, e.g. approaches for p-value adjustments in hypothesis testing (Abramowicz et al., 2018; Pini & Vantini, 2017b; Vsevolozhskaya et al., 2014) as well as simultaneous confidence bands based on results from random field theory (Liebl & Reimherr, 2020, 2023; Telschow & Schwartzman, 2022). The former approaches rely on permutation tests rather than distributional assumptions and are therefore computationally intensive. We here expand the methods and results obtained by Liebl and Reimherr (2020, 2023) to deal with a functional causal parameter estimated with an outcome regression estimator, and thereby obtain simultaneous confidence bands for the FATE. These simultaneous confidence bands control the probability that the true FATE lies outside the bands over its entire domain and thus

provide global control of the coverage probability. In addition, they are less computationally intensive than many simulation-based approaches, and their widths can adapt to the local structure of the data (Liebl & Reimherr, 2023).

The rest of the article is organised as follows. Section 2 presents the theory and methods for drawing exact finite sample inference using simultaneous confidence bands for the functional causal parameter of interest. Section 3 presents the study of the effect of early adult residence location on cumulative lifetime incomes using the methods introduced. Numerical experiments inspired from the real data application are used in Section 4 to evaluate the performance of the simultaneous confidence bands when the covariance function is estimated. Section 5 concludes the article with a discussion.

2 Theory: functional average treatment effect

2.1 Definition and identification

As mentioned above, our goal is to estimate the causal effect of early adult residence location on income accumulation over time for one Swedish birth cohort. This example will be used throughout the section for illustration.

To simplify notation, we do not distinguish between random variables and their realisations. Let z_i be a binary treatment indicator for individual i, taking the value $z_i = 1$ if individual i is exposed to treatment, and $z_i = 0$ otherwise. In the motivating example, this would correspond to living in an urban and a rural location, respectively, at the age of 20. Corresponding to these treatment levels, we define the two potential outcome processes $y_{1i} = \{y_{1i}(t){:}t \in [0, 1]\}$ and $y_{0i} = \{y_{0i}(t){:}t \in [0, 1]\}$, here cumulative incomes over time. These processes are assumed to be once continuously differentiable almost surely and without loss of generality, the domain of the argument t has been scaled to $[0, 1]$. For each individual, only one of these potential outcome processes can be observed, depending on whether they received the treatment or not. The actually observed outcome process is assumed to be $y_i(t) = y_{1i}(t)z_i + y_{0i}(t)(1 - z_i)$. This, together with the fact that treatment assignment for one unit does not affect other units' potential outcomes is often called the 'stable unit treatment value assumption (SUTVA)' (Imbens & Rubin, 2015, Chap 1). In our application, this entails that individuals' choice of residence location at age 20 does not affect other's lifetime income processes. While violation of such an assumption cannot be ruled out empirically, it is reasonable to assume that violations of SUTVA, i.e. interference between individuals, are negligible since the regions considered are large in area and population, and we focus only on one cohort. In addition, let $\boldsymbol{x}_i = (1, x_{1i}, \ldots, x_{Ki})^T$ be a vector of covariates measured prior to treatment assignment, i.e. they are assumed not to be affected by treatment. This vector includes a 1-element for ease of notation later on. We have a random sample of size n, with $i = 1, \ldots, n$, for which we observe $\boldsymbol{x}_i$, z_i and $y_i = \{y_i(t){:}t \in [0, 1]\}$. Let n_1 and $n_0 = n - n_1$ be the number of treated and untreated units in this sample, respectively.

The focus in this article is on one causal parameter, the Functional Average Treatment Effect (FATE):

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

An alternative causal parameter of interest is the Functional Average Treatment Effect on the Treated population (FATT). The developments for this parameter are given in Appendix A.

If the treatment assignment is randomised, the FATE can be identified directly from the observed data. However, this is not the case in our case study or in many other practical applications. Individuals who live in an urban area at the age of 20 could differ from those living in a rural area with respect to characteristics which in turn relate to income accumulation. In the presence of such confounding, identification of the causal effect from the observed data requires further assumptions (e.g. Holland, 1986; Rosenbaum & Rubin, 1983; Wooldridge, 2010). Here, we use the following identification assumptions:

Assumption 1 (Overlap).

$$\text{a. } \Pr(z_i = 0 \mid \boldsymbol{x}_i) > 0, \quad \forall\, \boldsymbol{x}_i,$$
$$\text{b. } \Pr(z_i = 1 \mid \boldsymbol{x}_i) > 0, \quad \forall\, \boldsymbol{x}_i.$$

Assumption 2 (Mean ignorability).

$$\text{a. } \mathbb{E}\big(y_{0i}(t) \mid z_i, \boldsymbol{x}_i\big) = \mathbb{E}\big(y_{0i}(t) \mid \boldsymbol{x}_i\big) \quad \forall\, t \in [0, 1],$$
$$\text{b. } \mathbb{E}\big(y_{1i}(t) \mid z_i, \boldsymbol{x}_i\big) = \mathbb{E}\big(y_{1i}(t) \mid \boldsymbol{x}_i\big) \quad \forall\, t \in [0, 1].$$

Mean ignorability (Wooldridge, 2010, Section 21.3) is a weaker version of the distribution ignorability assumption often used in the literature, where the potential outcomes and the treatment assignment are assumed to be independent in distribution conditional on the pre-treatment covariates. When an ignorability assumption holds, we say that $\boldsymbol{x}$ contains all confounding covariates.

The causal parameter θ is identified under the above assumptions since we can then write, for all $t \in [0, 1]$:

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

Assumption 1 implies that there are no combinations of values on the pre-treatment covariates for which an individual would have probability zero of living in either an urban or a rural area at the age of 20. This assumption can be checked empirically as we did with our case study, see Appendix D for details. Note that in any case this assumption can be enforced by discarding individuals from the population, changing thereby the population and parameter of interest. On the other hand, Assumption 2 cannot be empirically corroborated and there is always a possibility that some unobserved characteristic confounds the effect of individuals' residence location at 20 years on lifetime incomes. For instance, while we will control for educational attainments, such a covariate might not measure all innate skills that could affect both residence location and lifetime incomes. This is a common challenge when drawing causal inferences from observational data, and it means that the results should be interpreted with a certain level of caution.

2.2 Estimation and inference

We consider an outcome-regression estimator for θ, which assumes linear function-on-scalar models (Ramsay & Silverman, 2005) for the outcomes:

$$y_{zi}(t) = \boldsymbol{x}_i^T \boldsymbol{\beta}_z(t) + \varepsilon_{zi}(t), \quad z = 0, 1, \tag{1}$$

where $\boldsymbol{\beta}_z(t)$ are $K + 1$-dimensional functional parameters. Usual diagnostic tools can be used to include covariates that are transformations of the original covariates, although the inference will not then take into account such a model building step.

In order to obtain exact finite sample results below, we make the following assumptions on the random error processes $\varepsilon_{zi} = \{\varepsilon_{zi}(t) : t \in [0, 1]\}$.

Assumption 3 (Random error processes). The vector $\boldsymbol{\varepsilon}_z = (\varepsilon_{z1}, \ldots, \varepsilon_{zn})^T$ of error terms in model (1) is a multivariate Gaussian process:

$$\boldsymbol{\varepsilon}_z \sim \mathcal{MGP}\big(0, \sigma_z, \mathbb{I}_{n_z}\big),$$

where $\sigma_z(s, t) = Cov(\varepsilon_{zi}(s), \varepsilon_{zi}(t))$, $s, t \in [0, 1]$, and $\mathbb{I}_{n_z}$ is a $n_z \times n_z$ identity matrix.

Let $\boldsymbol{y}(t) = (y_1(t), \ldots, y_n(t))^T$ denote the vector of observed outcomes, with $\boldsymbol{y}_1(t) = \{y_i(t) : z_i = 1\}$ and $\boldsymbol{y}_0(t) = \{y_i(t) : z_i = 0\}$ its subsets for the treated and untreated units, respectively. Similarly, let $\boldsymbol{X}_1$ be the $n_1 \times (K + 1)$ matrix containing observations on the K covariates (and a column of 1 for the intercept) for the treated subgroup, and $\mathbf{X}_0$ the corresponding $n_0 \times (K + 1)$ matrix of

covariates for the non-treated, i.e. subsets of $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$. Then, the functional regression coefficients can be estimated by using ordinary least squares:

$$\hat{\boldsymbol{\beta}}_z(t) = (\mathbf{X}_z^T \mathbf{X}_z)^{-1} \mathbf{X}_z^T \boldsymbol{y}_z(t), \tag{2}$$

$\hat{\boldsymbol{\beta}}_z = \{\hat{\boldsymbol{\beta}}_z(t) : t \in [0, 1]\}$.

We can then construct an estimator for the FATE by imputing the unobserved potential outcomes with predictions from the fitted outcome models above:

$$\begin{aligned} \hat{\theta}(t) &= \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_1(t) - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_0(t)) \\ &= \overline{\boldsymbol{x}} \hat{\boldsymbol{\beta}}_1(t) - \overline{\boldsymbol{x}} \hat{\boldsymbol{\beta}}_0(t), \end{aligned} \tag{3}$$

where $\{\overline{\boldsymbol{x}} = \frac{1}{n} \mathbb{1}_{1 \times n} \mathbf{X}$ is the vector of sample means. This yields the estimator $\hat{\theta} = \{\hat{\theta}(t) : t \in [0, 1]\}$.

We now give the finite sample properties of the estimator $\hat{\theta}$. Throughout, we assume a fixed design for the regression models in (1), meaning that inference is conditional on $\mathbf{X}$.

Theorem 1 Let SUTVA, and Assumptions 1–3 hold. Then, the estimator $\hat{\theta}$ is a Gaussian process:

$$\hat{\theta} \mid \mathbf{X} \sim \mathcal{GP}\left(\overline{\boldsymbol{x}} \boldsymbol{\beta}_1 - \overline{\boldsymbol{x}} \boldsymbol{\beta}_0, C_\theta\right), \tag{4}$$

where

$$C_\theta(s, t) = \sigma_1(s, t) \overline{\boldsymbol{x}} \left(\mathbf{X}_1^T \mathbf{X}_1\right)^{-1} \overline{\boldsymbol{x}}^T + \sigma_0(s, t) \overline{\boldsymbol{x}} \left(\mathbf{X}_0^T \mathbf{X}_0\right)^{-1} \overline{\boldsymbol{x}}^T. \tag{5}$$

See Appendix B for a proof. Note that the Gaussian condition in Assumption 3 is not needed to derive well known pointwise properties of $\hat{\theta}(t)$ such as consistency for $\theta(t)$ (e.g. Genbäck & de Luna, 2019). We use this assumption to obtain exact finite sample inference on the functional estimator.

Using Theorem 1, we can build on the simultaneous confidence bands presented in Liebl and Reimherr (2020, 2023), which allow for a global control of the coverage probability over the entire domain, to quantify the uncertainty in the estimation of θ. A $(1 - \alpha)\%$ simultaneous confidence band for θ is given by

$$\hat{\theta}(t) \pm u_\alpha(t) \sqrt{C_\theta(t, t)}, \tag{6}$$

where $u_\alpha = \{u_\alpha(t) : t \in [0, 1]\}$ is a continuous threshold function such that, for a given level α,

$$P\left(\theta(t) \in \hat{\theta}(t) \pm u_\alpha(t) \sqrt{C_\theta(t, t)} \quad \forall\, t \in [0, 1]\right) \geq 1 - \alpha.$$

In other words, (6) yields a $(1 - \alpha)$ simultaneous confidence band for the functional parameter θ, which is such that $\theta(t)$ lies within the band over the entire domain with a probability of at least $(1 - \alpha)$. See Liebl and Reimherr (2023) for details on how these bands are constructed. If we were to instead use a naive constant threshold of, e.g. $u_{0.05}(t) = 1.96$ for all t, this would yield simultaneous coverage lower than 0.95 due to the multiple comparison problem mentioned earlier. In addition to the aforementioned global control of the coverage probability, a unique feature of the simultaneous confidence bands is that they also allow to draw local inference, e.g. giving at which significance level one can reject the null hypothesis of no effect within a sub-interval of $[0, 1]$ (Liebl & Reimherr, 2023, Proposition 3.2).

In practice, the covariance structure of the random errors functions (and in extension, C_θ) is likely unknown. However, we can use the residuals from the two outcome regression fits as estimates of the error functions, and their sample covariance function as a consistent estimate $\hat{\sigma}_z(s, t)$ of $\sigma_z(s, t)$ to obtain:

$$\hat{C}_\theta(s, t) = \hat{\sigma}_1(s, t)\ \overline{\boldsymbol{x}}\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\overline{\boldsymbol{x}}^T + \hat{\sigma}_0(s, t)\overline{\boldsymbol{x}}\left(\mathbf{X}_0^T\mathbf{X}_0\right)^{-1}\overline{\boldsymbol{x}}^T. \tag{7}$$

When this estimate is used in finding $u_\alpha(t)$,a t-process can be used as a finite sample correction, see Liebl and Reimherr (2023).

3 The effects of early adult residence location on cumulative lifetime income

We aim at studying the consequences of the place of labour market entry on lifetime income trajectories. We focus our analysis on the cohort born 1954 in Sweden, using a population-wide and longitudinal database available at the Umeå SIMSAM Lab (Lindgren et al., 2016). This database allows us to follow income trajectories from the cohort's entrance into the labour market until 2017 (age 63). The potential effect of the time (and place) of labour market entrance on incomes has been widely discussed in the Economics literature; see, e.g. Altonji et al. (2016), Åslund and Rooth (2007), Kwon et al. (2010), Oreopoulos et al. (2012), Raaum and Røed (2006), and Schwandt and von Wachter (2019). We here follow an operationalisation of (exposure to) treatment similar to Raaum and Røed (2006) and define the location of the entry in the labour market as the place of residence at the age of 20, i.e. in 1974 for our cohort. Thus, we define a binary treatment, where $z = 1$ is defined as living at the age of 20 in the local labour markets of the three biggest Swedish cities (Stockholm, Gothenburg, or Malmö). The alternative treatment, $z = 0$, is defined as living in local labour markets in which the central municipality had less than 50,000 residents at the age of 20. Local labour markets are units defined by patterns of employment and commuting at the municipality level (Karlsson & Olsson, 2006; Statistics Sweden, 2010).

The outcome of interest, $y(t)$, is the total earned income (in SEK), defined as all taxable income except from capital. This covers wages and income from business/self-employment, unemployment, parental and sick leave benefits, and pensions. The outcome is measured annually, and data are available between 1968 and 2017. The income is adjusted for inflation to match the monetary value in 2017 and discounted with a factor of 0.03 (Johansson & Kriström, 2016). We calculate the cumulative income over the period 1975–2017, i.e. starting from the year following treatment exposure. Lastly, this cumulative income is logarithmised, a common transformation for incomes that normalises their distribution. In the sequel, 'income' will refer to the logarithmised cumulative income.

We control for a number of pre-treatment covariates: *Year of first income*, the year in which an individual first received income if this occurs before exposure to the treatment (before 1974); *Previous income*, which is the logarithmised cumulative income an individual has accumulated before exposure; *Children before age 20*, a binary indicator of the subject having at least one child before exposure; and an indicator of completing (at least) upper *secondary education*. The latter is based on the subjects' highest achieved education level in 1990 as a proxy, as this is the first year for which the variable is available annually. In addition, we control for some family measures: The subjects' *number of siblings* (in 1973); *Parents' income* during the subjects' adolescence (the logarithmised total for both parents for the years 1968–1971); and an indicator indicating if both parents are born outside of Sweden (*Foreign-born parents*). We also control for the highest education level (lower secondary, upper secondary, or tertiary) achieved by either mother, father, or both, in 1970, measured by the indicators of *parents' secondary education* and *parents' tertiary education*. Descriptive statistics for the confounders are shown in Table D.1 in Appendix D.

The 1954 birth cohort was chosen to give the longest possible period in which income data is available. We limit the analysis to individuals born in Sweden and excluded a total of around 4,000 individuals due to missingness on one or more of the covariates, primarily those related to parental background. The resulting data consist of 54,485 individuals: 27,805 men (55.6% 'exposed', $z = 1$) and 26,680 women (59.3% 'exposed', $z = 1$). In our analyses, we treat men and women as separate populations and target two parameters of interest: the FATE for men, θ_M,

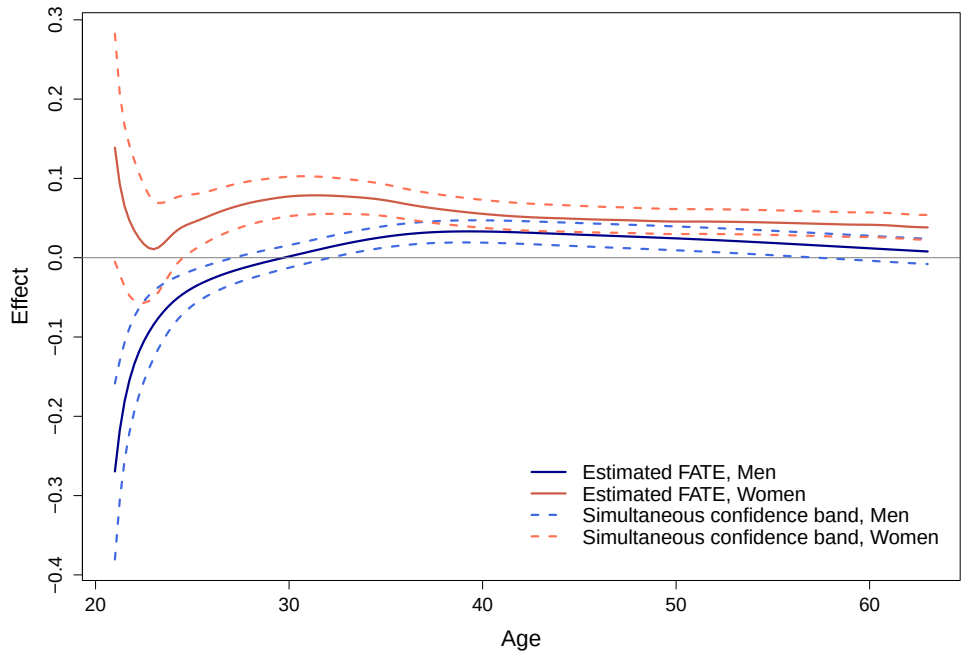

Figure 2. Estimates (solid lines) of the FATE of living in urban versus rural areas at the age of 20 on cumulative life income for the cohort born 1954 in Sweden. 95% simultaneous confidence bands (dashed lines) for the FATE.

and for women, θ_W. After some initial pre-processing, we estimate θ_M and θ_W with (3) and construct 95% confidence bands as described in Section 2.2. Appendix C contains further details of this process. All analyses are performed in `R` (Version 4.0.2) (R Core Team, 2020), using the `ffscb` package (Liebl & Reimherr, 2019) to obtain the confidence bands.

Figure 2 shows the estimates of the two FATE θ_W and θ_M and their corresponding 95% simultaneous confidence bands from the separate analyses in the populations of men (blue, lower set of curves) and women (red, upper set of curves). Especially in the early years, the two effects display different patterns between the two populations. A caveat to this comparability and to the results discussion below is the possible existence of unobserved confounding. For men, we observe a negative effect of living in an urban area at the age of 20 on early incomes, between the ages of 21 and 26 (1975–1980). This might be due to differences in later educational attainment, i.e. that those who live in an urban area in early adulthood are more likely to continue tertiary education. From around age 30 and onward, the effect for men is positive, and the 95% simultaneous confidence bands do not contain zero for the majority of these years.

For women, the effect is positive for the entire period, and the bands do not contain zero from 1979 onward. This suggests that women already at an early stage in adulthood tend to benefit from living in an urban location at age 20 and continue to do so for most of their working life. Notice from Figure 2 that in the first years after (potential) labour market entry, the confidence bands are wider to account for the larger variability in cumulative incomes during that period. Additional figures comparing the simultaneous confidence bands to pointwise bands can be found in Appendix D.

4 Simulation study

This section evaluates the finite sample performance of the confidence bands introduced in this article using Monte Carlo simulations. A realistic setting is obtained by sampling covariate values from the sample used in the case study of Section 3 and simulating exposure and outcome using the fitted propensity score and outcome models.

4.1 Study design

The design described below is replicated for the women and men samples separately. We sample n vector values, $\boldsymbol{x}_i$, $i = 1, \ldots, n$, with replacement from the sample used in Section 3 to form a simulation replicate of the original sample. We use stratified sampling since some covariate value combinations have few observations. The three strata used are 'Children before age 20', 'Foreign-born parents', and 'Parents' tertiary education'. These simulated values of $\boldsymbol{x}$ are treated as fixed for all replicates in a fixed design setting. A random design setting will also be considered later on.

We generate the treatment assignments as $z_i \sim \text{Ber}(\pi_i)$, with

$$\pi_i = \Pr(z_i = 1 \mid \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i^T \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{x}_i^T \boldsymbol{\gamma})}, \tag{8}$$

where $\boldsymbol{\gamma}$ is based on estimates from the application study and given in Table E.1 in Appendix E. The potential outcomes are generated at 43 equidistant points t on [0,1] as

$$y_i(t) = \begin{cases} \boldsymbol{x}_i^T \boldsymbol{\beta}_0(t) + \varepsilon_{0i}(t), & \text{for } i{:}z_i = 0, \\ \boldsymbol{x}_i^T \boldsymbol{\beta}_1(t) + \varepsilon_{1i}(t), & \text{for } i{:}z_i = 1, \end{cases} \tag{9}$$

where $\boldsymbol{\beta}_z(t)$, $z = 0, 1$, are based on the estimated regression coefficients from the corresponding models in the application study. They are displayed in Figure E.1 in the Appendix (exact numerical values can be obtained from the authors). Moreover, the random error functions $\varepsilon_{zi}(t)$ are generated such that

$$\varepsilon_{zi} \sim \mathcal{GP}(0, \sigma_z), \tag{10}$$

where σ_z are calculated from the residuals of the outcome regression models in Section 3. In the fixed design setting, the '*true*' FATE parameter is $\{\bar{\boldsymbol{x}}^T \boldsymbol{\beta}_1(t) - \{\bar{\boldsymbol{x}}^T \boldsymbol{\beta}_0(t)$, where $\{\bar{\boldsymbol{x}} = \frac{1}{n} \boldsymbol{x}^T \mathbb{1}_{n \times 1}$.

We use 1,000 replicates to evaluate four different types of pointwise and simultaneous confidence bands for $\hat{\theta}(t)$:

(i) We assume that the covariance structures σ_0 and σ_1, used to generate the random error functions in (10), are known. These covariances are used to calculate $C_\theta(s, t)$. We then construct 95% simultaneous confidence bands as in (6).

(ii) We estimate the error functions' covariance structures σ_0 and σ_1 using the residuals from the regression models in the current simulation replicate. This estimated covariance structure is then used to construct the confidence bands.

(iii) We construct pointwise 95% confidence intervals separately at each of the 43 time points for which the outcome is observed. For this third type of band, we use the diagonal of the known true covariance matrix $C_\theta(s, t)$ to construct the pointwise 95% confidence band:

$$\hat{\theta}(t) \pm t_{0.05} \sqrt{C_\theta(t, t)}, \tag{11}$$

where $t_{0.05}$ is the 97.5th percentile from the t-distribution with $n - 1$ degrees of freedom.

(iv) We construct the pointwise confidence bands with (11), but using the estimator $\hat{C}_\theta(t, t)$ in (7) instead of $C_\theta(t, t)$.

The empirical coverage rate is computed as 100 times the proportion of replicates for which the band contains $\theta(t)$ over all observed time points t. Four different sample sizes are used, $n =$ 250, 500, 1,000 and 10, 000.

Aside from the above setup using a fixed design and Gaussian random errors, we also compare the bands of type (ii) and (iv) in two other setups: Firstly, we perform the simulations in a random design setup, in which new covariate values are re-sampled for each replicate. For the evaluation of

Table 1. Empirical coverages based on 1,000 replicates for four types of confidence bands in the fixed design, four sample sizes, presented for men and women separately

Fixed design	Men				Women			
Type\n	250	500	1,000	10,000	250	500	1,000	10,000
(i)	95.0	96.4	95.7	95.2	95.2	96.0	95.9	95.7
(ii)	93.4	96.1	95.5	95.1	93.2	95.4	95.8	95.6
(iii)	70.6	69.7	70.9	72.1	71.0	71.2	72.7	70.1
(iv)	66.2	67.5	69.7	72.1	65.5	65.5	70.8	69.9

Note. The nominal coverage level is 95%.

the confidence bands, we use the approximation of the 'true' FATE parameter:

$$\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}\left(\bar{\boldsymbol{x}}_b^T\boldsymbol{\beta}_1(t)-\bar{\boldsymbol{x}}_b^T\boldsymbol{\beta}_0(t)\right),$$

where $\boldsymbol{x}_b$ is the covariate values in the bth replicate.

Secondly, for the fixed design setup, we allow in another scenario for an error process which is not Gaussian: instead of (10), we generate the error term from a multivariate t-distribution with ten degrees of freedom (Genz et al., 2020): $\varepsilon_{zi}(t) \sim t_{10}(\mathbf{0}, \sigma_z(t, t))$.

4.2 Results

Table 1 displays empirical coverage rates for the four confidence bands studied obtained over the 1,000 replicates in the fixed design setup. There are no qualitative differences in the results between the simulations based on the data for men and women. The simultaneous confidence bands using the known covariance [type (i)] attain their nominal size of 95%. When estimating the covariance function $\hat{C}_\theta(s, t)$ [bands of type (ii)], we see that empirical coverages attain their nominal level of 95% for sample sizes $n = 500$ and larger. As expected, both types of pointwise confidence bands [type (iii) and (iv)] have too low empirical coverages due to the multiple comparison problem.

Figure 3 displays the results graphically from one simulation replicate with $n = 500$ using the fixed design and the male subset. The simultaneous confidence bands [type (i) and (ii)], shown as dotted and dashed red curves are wider than their pointwise counterparts and cover the true $\theta(t)$ at all points. An additional figure displaying the average estimates and confidence bands over these 1,000 replicates can be found in Appendix E.2.

The results for the two additional scenarios can be found in Table 2. We can see that the results from the random design setting are similar to those from the fixed design. The bands attain a coverage close to the nominal level of 95% for the sample sizes considered. In addition, the simultaneous confidence bands also perform well under non-Gaussian random error terms (see rows corresponding to Non-GP errors in Table 2).

5 Discussion

This article presents methods to study causal effects of a treatment on a functional outcome. In this context, we define a functional causal parameter (FATE or FATT), propose an estimator based on outcome regression, and provide simultaneous confidence bands that cover the true functional parameter over its entire domain with a given probability. The theoretical results are confirmed in simulation experiments, where the simultaneous confidence bands achieve the desired nominal coverage rate.

The methods derived here have allowed us to study the effect of early adult residence location on cumulative lifetime incomes over the entire working life of a Swedish birth cohort. For men, we find a negative effect of living in an urban versus a rural area at the age of 20 on cumulative

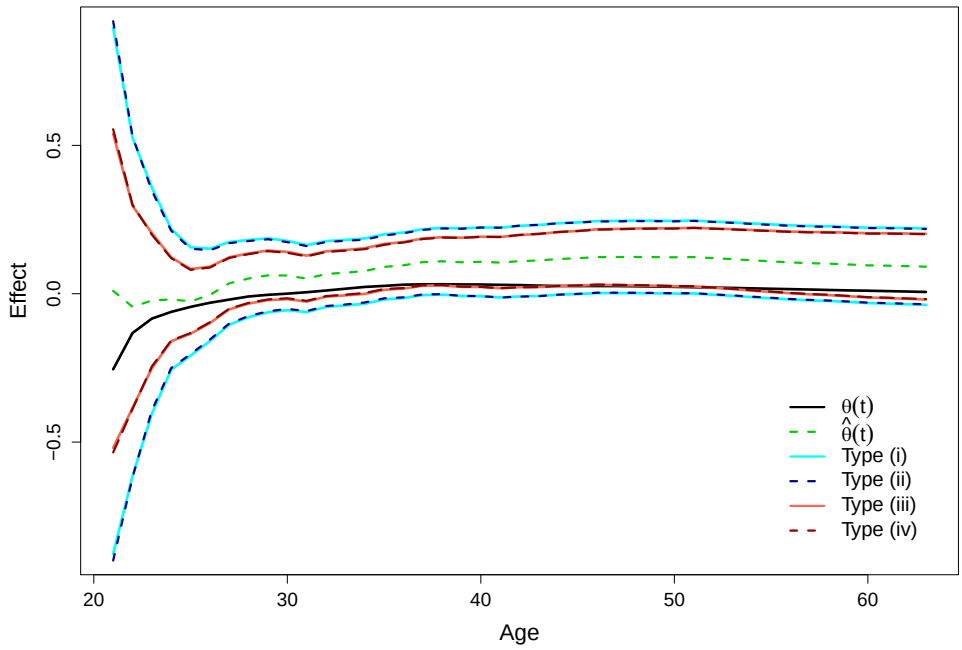

Figure 3. True and estimated FATE and the four types of confidence bands from one simulated replicate for the fixed design and male subset, $n = 500$.

Table 2. Empirical coverages based on 1,000 replicates for additional scenarios, presented for men and women separately

Setup	Type\n	Men				Women			
		250	500	1,000	10,000	250	500	1,000	10,000
Random design	(ii)	93.8	95.3	94.6	96.9	95.5	94.6	95.6	96.3
	(iv)	67.3	70.3	72.4	72.9	67.0	67.3	69.0	69.9
Non-GP errors	(ii)	93.4	95.3	96.0	95.2	95.2	96.2	95.4	96.1
	(iv)	66.9	69.2	70.9	71.5	63.9	67.3	68.0	72.3

Note. The nominal coverage level is 95%.

incomes in early career stages. This effect becomes positive later in life. For women, living in an urban area in early adulthood has a slightly stronger, positive effect on cumulative incomes throughout the working life. The results of such an observational study must be taken with caution, in particular because of the possibility of existing unobserved confounding (Assumption 2 violated). On the other hand, earlier observational studies on the associations between initial labour market conditions and later labour market outcomes have used much shorter follow-up periods (Altonji et al., 2016; Åslund & Rooth, 2007; Kwon et al., 2010; Oreopoulos et al., 2012; Schwandt & von Wachter, 2019), or do not correct for the multiple comparison problem (Raaum & Røed, 2006). Note that the methods and theory presented are readily applicable to other contexts. For instance, patients may be monitored with electronic instruments yielding functional signal outcomes in medical applications.

We have focused on outcome regression estimators and only allow for scalar covariates. Theoretical results for other estimators of average causal effects, including the augmented inverse probability weighting estimator and the closely related targeted learning estimator (van der Laan & Rose, 2018), as well as to allow for functional covariates would be valuable future research

directions. In all the above cases, the difficulty resides in showing that the resulting functional estimator is asymptotically a Gaussian or elliptical process. In particular, the tightness of the pointwise estimators needs to be shown, which is known to be problematic with functional covariates (Choi & Reimherr, 2018).

Acknowledgments

We are grateful to Xijia Liu for comments that have improved the article.

Conflict of interests: The authors declare no competing interests.

Funding

The work was supported by the Swedish Research council [grant number 2016-02851]. The Umeå SIMSAM Lab data infrastructure in Umeå used in this study was developed with support from the Swedish Research Council [grant number 2008-7491], the Riksbanken Jubileumsfond and by strategic funds from Umeå University.

Data availability

The data analysis is based on a record linked register database available at the Umeå SIMSAM Laboratory (https://www.umu.se/en/research/infrastructure/umea-simsam-lab/) at Umeå University, Sweden. The database is built as a combination of different population-based registers linked through Swedish personal registration numbers and was compiled in collaboration with different Swedish authorities. Both the approval from the Ethical vetting board and the contracts we have signed with the Swedish authorities do not allow us to give away the data to a third party. The data can, however, be accessed by researchers wanting to replicate the analysis, although this can be done only locally at the Umeå SIMSAM Laboratory, where the data is stored on servers disconnected from the internet.

Appendix A. Functional average treatment effect on the treated population

The Functional Average Treatment effect on the Treated population (FATT) is defined as:

$$\eta := \left\{\eta(t) = \mathbb{E}\big(y_{1i}(t) - y_{0i}(t) \mid z_i = 1\big) : t \in [0, 1]\right\}.$$

It is identified under Assumptions 1a and 2a in Section 2.1. Under model (1), an estimator for the FATT is given by:

$$\begin{aligned}\hat{\eta}(t) &= \frac{1}{n_1}\sum_{i=1}^{n} z_i(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_1(t) - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_0(t)) \\ &= \overline{\boldsymbol{x}}_1\hat{\boldsymbol{\beta}}_1(t) - \overline{\boldsymbol{x}}_1\hat{\boldsymbol{\beta}}_0(t),\end{aligned}$$

where $\{\bar{x}_1 = \frac{1}{n_1}\mathbb{1}_{1\times n_1}\mathbf{X}_1$ is the vector of sample means for the subset of treated observations. Analogous to equations (4)–(5), we have that $\hat{\eta} = \{\hat{\eta}(t) : t \in [0, 1]\}$ is a Gaussian process:

$$\hat{\eta} \sim \mathcal{GP}(\overline{\boldsymbol{x}}_1\boldsymbol{\beta}_1 - \overline{\boldsymbol{x}}_1\boldsymbol{\beta}_0, C_\eta),$$

where

$$C_\eta(s, t) = \sigma_1(s, t)\overline{\boldsymbol{x}}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\overline{\boldsymbol{x}}_1^T + \sigma_0(s, t)\ \overline{\boldsymbol{x}}_1\left(\mathbf{X}_0^T\mathbf{X}_0\right)^{-1}\overline{\boldsymbol{x}}_1^T.$$

Appendix B. Estimator of FATE is a Gaussian process

We reproduce here the definition of a multivariate Gaussian process given in Chen et al. (2020):

Let $\boldsymbol{f}$ be a multivariate Gaussian process on $\mathcal{X}$ with vector-valued mean function $\boldsymbol{\mu}$:$\mathcal{X} \mapsto \mathbb{R}^n$, covariance function σ:$\mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, and positive semi-definite parameter matrix $\Omega \in \mathbb{R}^{n \times n}$. We denote this process $\boldsymbol{f} \sim \mathcal{MGP}(\boldsymbol{\mu}, \sigma, \Omega)$. Then, the vectorisation of any finite collection of L vector-valued variables have a joint multi-variate Gaussian distribution:

$$vec\Big([\boldsymbol{f}(x_1)^T, \ldots, \boldsymbol{f}(x_L)^T]^T\Big) \sim \mathcal{N}_{dn}\big(vec(M), \Sigma \otimes \Omega\big), L \in \mathbb{N},$$

where $M \in \mathbb{R}^{n \times L}$ with $M_{ij} = \mu_j(x_i)$ and $\Sigma > 0$, $\Sigma \in \mathbb{R}^{d \times d}$ with $\Sigma_{ij} = \sigma(x_i, x_j)$. Furthermore, $\boldsymbol{f}, \boldsymbol{\mu} \in \mathbb{R}^n$ are row vectors whose components are the functions $\{f_i\}_{i=1}^n$ and $\{\mu_i\}_{i=1}^n$, respectively. Sometimes Σ is called the column covariance matrix, while Ω is the row covariance matrix. See also Chen et al. (2021) for proof of existence and further properties. This in turn implies that any finite collection of L vector-valued variables have a joint matrix variate Gaussian distribution (Gupta & Nagar, 2000):

$$[\boldsymbol{f}(x_1)^T, \ldots, \boldsymbol{f}(x_L)^T]^T \sim \mathcal{N}_{d,n}(M, \Sigma \otimes \Omega).$$

Proof of Theorem 1: Assumption 3 states that the error terms in model (1) are multivariate Gaussian processes, denoted as

$$\boldsymbol{\varepsilon}_z \sim \mathcal{MGP}(0, \sigma_z, \mathbb{I}_{n_z}),$$

where $\mathbb{I}_{n_z}$ is a $n_z \times n_z$ identity matrix. Using the definition above we have that, for any finite collection $\{t_l\}_{l=1}^L$ of points in the domain, the error terms for each treatment group follow a joint matrix-variate Gaussian distribution:

$$\mathrm{E}_z := \Big[\boldsymbol{\varepsilon}_z(t_1)^T, \ldots, \boldsymbol{\varepsilon}_z(t_L)^T\Big] \sim \mathcal{N}_{n_z,L}\big(0_{n_z,L}, \mathbb{I}_{n_z} \otimes \Sigma_z\big),$$

where $\boldsymbol{\varepsilon}_z(t_l)$, $l = 1, \ldots, L$, are row vectors of length n_z; $0_{n_z,L}$ is a $n_z \times L$ null matrix, and Σ_z is a $L \times L$ matrix with elements $[\Sigma_z]_{i,j} = \sigma_z(t_i, t_j)$.

We further specify the following matrices: $\Omega_z := (\mathbf{X}_z^T \mathbf{X}_z)^{-1}$, $\mathrm{B}_z := [\boldsymbol{\beta}_z(t_1), \ldots, \boldsymbol{\beta}_z(t_L)]$, and $\hat{\mathrm{B}}_z := [\hat{\boldsymbol{\beta}}_z(t_1), \ldots, \hat{\boldsymbol{\beta}}_z(t_L)]$. Given a fixed design matrix, the matrix of estimated regression coefficients can be expressed as

$$\hat{\mathrm{B}}_z = \mathrm{B}_z + \Omega_z \mathbf{X}_z^T \mathrm{E}_z.$$

Then, for any finite collection of points, $\hat{\mathrm{B}}_z$ follows a matrix-variate Gaussian distribution (Gupta & Nagar, 2000) with

$$\hat{\mathrm{B}}_z \sim \mathcal{N}_{(K+1),L}\Big(\Omega_z \mathbf{X}_z^T 0_{n_z,L} + \mathrm{B}_z, \Big[\Omega_z \mathbf{X}_z^T \mathbb{I}_{n_z} \big(\Omega_z \mathbf{X}_z^T\big)^T\Big] \otimes \Sigma_z\Big),$$

which simplifies to

$$(B1)\hat{\mathrm{B}}_z \sim \mathcal{N}_{(K+1),L}(\mathrm{B}_z, \Omega_z \otimes \Sigma_z).$$

Because the above statement holds for any finite collection $\{t_l\}_{l=1}^L$,

the process $\hat{\boldsymbol{\beta}}_z = \{\hat{\boldsymbol{\beta}}_z(t){:}t \in [0, 1]\}$ is a multivariate Gaussian process:

$$\hat{\boldsymbol{\beta}}_z \sim \mathcal{MGP}(\boldsymbol{\beta}_z, \sigma_z, (\mathbf{X}_z^T\mathbf{X}_z)^{-1}).$$

To derive the distribution of $\hat{\theta}$, we can use the distribution of $\hat{\mathrm{B}}_z$ in (B1) and see that the product

$$\bar{\boldsymbol{x}}\hat{\mathrm{B}}_z \sim \mathcal{N}_{1,L}(\bar{\boldsymbol{x}}\mathrm{B}_z, [\bar{\boldsymbol{x}}\Omega_z\bar{\boldsymbol{x}}^T] \otimes \Sigma_z).$$

This $1 \times L$-dimensional matrix-variate Gaussian distribution coincides with a L-dimensional multivariate Gaussian distribution. Since these distributional statements, again, hold for any finite collection of points in the domain, it follows that $\{\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_z = \{\{\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_z(t){:}t \in [0, 1]\}$ is a Gaussian process:

$$\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_z \sim \mathcal{GP}(\bar{\boldsymbol{x}}\boldsymbol{\beta}_z, \bar{\boldsymbol{x}}\left(\mathbf{X}_z^T\mathbf{X}_z\right)^{-1}\bar{\boldsymbol{x}}^T\sigma_z).$$

Under the assumption of mean ignorability, $\{\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_0$ and $\{\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_1$ are independent, meaning that the difference between them also is a Gaussian process:

$$\hat{\theta} \sim \mathcal{GP}(\bar{\boldsymbol{x}}\boldsymbol{\beta}_1 - \bar{\boldsymbol{x}}\boldsymbol{\beta}_0, C_\theta),$$

with C_θ given in (5). □

Appendix C. Practical details of estimating the FATE and constructing confidence bands

The following section describes the technical details of the application study in section 3. All steps were performed in `R` (R Core Team, 2020), and done separately for men and women.

(1) To represent the discrete annual measurements of income as functional objects, we apply a low degree of monotone smoothing to the logarithmised cumulative incomes. The smoothing is performed using the `smooth.monotone` function in the `fda` package (Ramsay et al., 2020), and we use a basis of 45 cubic B-splines and a smoothing parameter of $\lambda = 0.1$, penalising the second derivative. During this stage, an additional 28 individuals had to be excluded from the analysis, because the shape of their income accumulations (very steep inclines followed by long plateaus) did not allow for a low degree of monotone smoothing. The resulting functional data objects are evaluated at a quarterly basis, i.e. yielding four evaluations per calendar year. Note that performing the same analyses of the data without or little smoothing yields very similar results.
(2) Next, we subset the observed data into two groups: treated ($z = 1$) and controls ($z = 0$).
(3) In each treatment group, we regress the quarterly evaluations of the smoothed functions, $y(t)$, on the covariates $\boldsymbol{x}$, using the `IWTlm` function in the `fdatest` package (Pini & Vantini, 2017a). The estimated regression coefficients $\hat{\boldsymbol{\beta}}_z(t)$ and the residuals from the two outcome regression models are saved.
(4) We then use the estimated regression coefficients and the covariate means to estimate the ATE with $\hat{\theta}(t)$ given in (3).
(5) We compute the sample covariances of the residuals from the two outcome regression models in step 2; yielding $\hat{\sigma}_1(s, t)$ and $\hat{\sigma}_0(s, t)$. These are used to compute $\hat{C}_\theta$ in (7).
(6) With $\hat{C}_\theta$, we use the `cov2tau_fun` function in the `ffscb` package (Liebl & Reimherr, 2019) to obtain the estimated roughness parameter function $\hat{\tau}$; see Liebl and Reimherr (2023).

(7) In the same package, we use the `confidence_bands` function to construct the simultaneous confidence bands using $\hat{\theta}$, $\hat{C}_\theta$ and $\hat{\tau}$. We choose to obtain 95% simultaneous confidence bands for $\theta(t)$ as well as the corresponding naive point-wise confidence bands based on the t-distribution. All other arguments of this function are set to their defaults.

Appendix D. Details of the case study in Section 3

D.1 Descriptives on the covariates

Table D.1 shows the distribution of the covariates for men and women separately.

Table D.1. Descriptive information on the covariates, presented as mean (standard deviation) or as % for indicator variables

Variable	Men	Women
Year of first income	1971.7 (1.2)	1972.1 (1.0)
Previous income	8.3 (4.1)	7.2 (4.5)
Children before age 20	1.9%	10.0%
Secondary education	71.0%	79.0%
Number of siblings	2.3 (1.8)	2.3 (1.8)
Parents' income	11.8 (0.9)	11.9 (0.9)
Foreign-born parents	2.3%	2.3%
Parents' secondary education	29.7%	30.0%
Parents' tertiary education	11.2%	11.3%

D.2 Empirical check of the overlap assumption

Figures D.1 and D.2 can be used to investigate Assumption 2 (Overlap) for men and women, respectively. For men, the support of the estimated propensity score ranges from 0.02 to 0.94 among the treated and from 0.01 to 0.91 among the controls. For women, the support of the estimated propensity score ranges from 0.02 to 0.94 among the treated and from 0.03 to 0.92 among the controls. This means that there is no evidence against the overlap assumption.

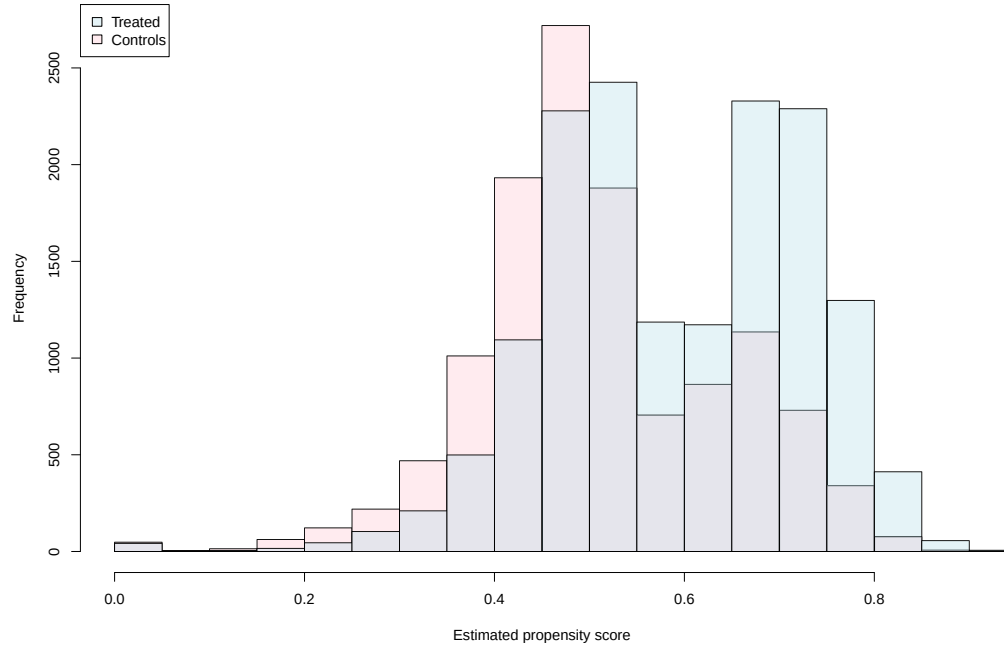

Figure D.1. Histogram of estimated propensity scores for the treated and controls among men.

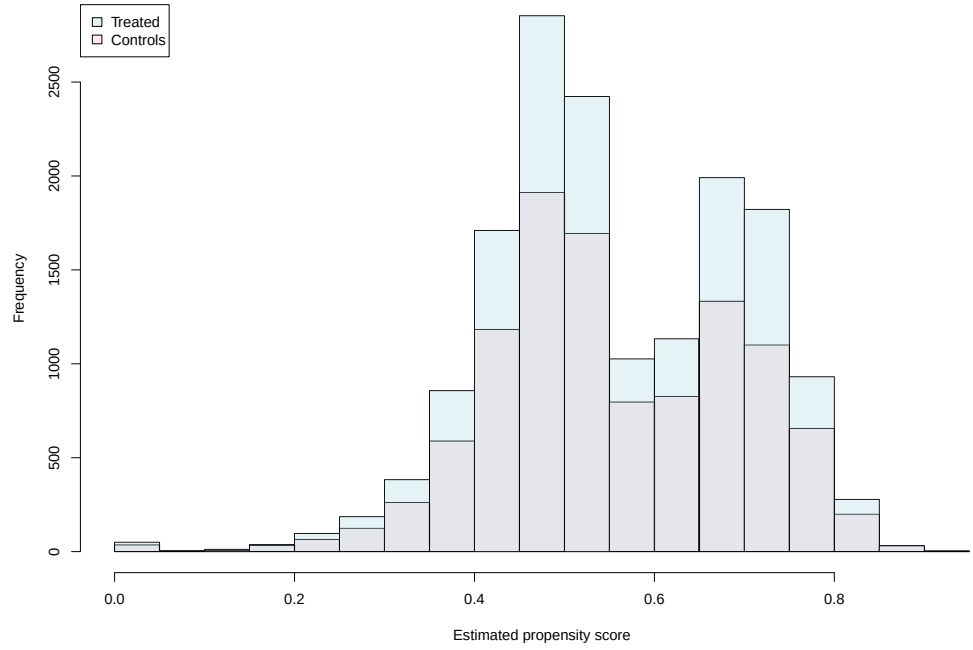

Figure D.2. Histogram of estimated propensity scores for the treated and controls among women.

D.3 Average estimates and confidence bands

Figures D.3 and D.4 display 95% pointwise confidence bands for the FATE estimated in Section 3, in addition to the simultaneous bands, for men and women respectively. We can see that the pointwise bands are more narrow, most notably so for women in the first 10 years.

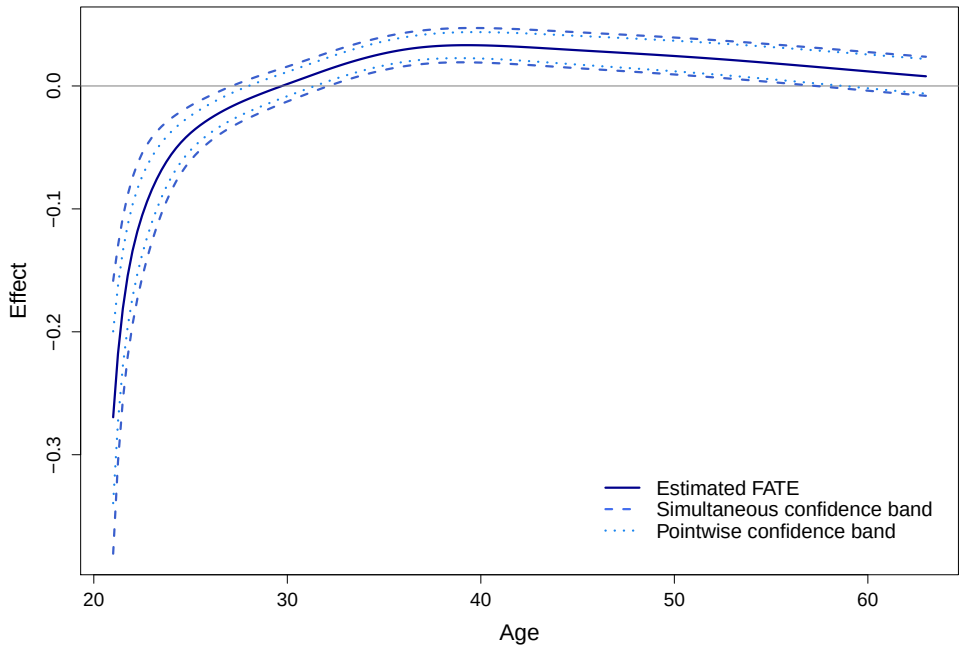

Figure D.3. Estimates (solid line) of the FATE of living in urban versus rural areas at the age of 20 on cumulative life income for the male cohort born 1954 in Sweden. 95% simultaneous confidence bands (dashed line) and 95% pointwise confidence bands (dotted line) for the FATE.

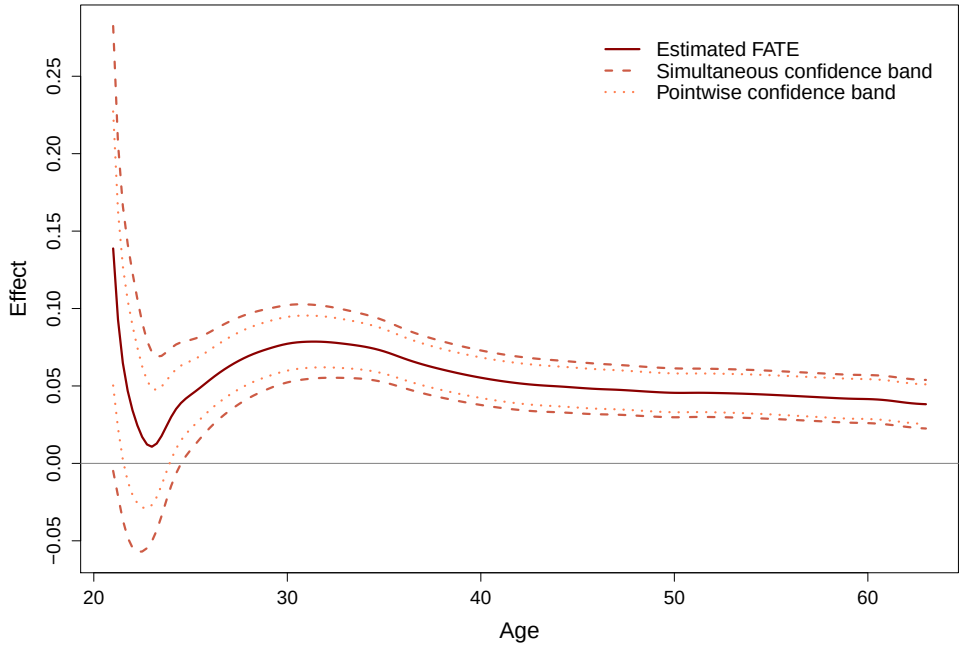

Figure D.4. Estimates (solid line) of the FATE of living in urban versus rural areas at the age of 20 on cumulative life income for the female cohort born 1954 in Sweden. 95% simultaneous confidence bands (dashed line) and 95% pointwise confidence bands (dotted line) for the FATE.

Appendix E. Details of the simulation study in Section 4

E.1 Parameters used to generate data

The coefficients used in the outcome models (9) are displayed in Figure E.1; exact values can be obtained from the authors. The coefficients used in the propensity score models in (8) are shown in Table E.1.

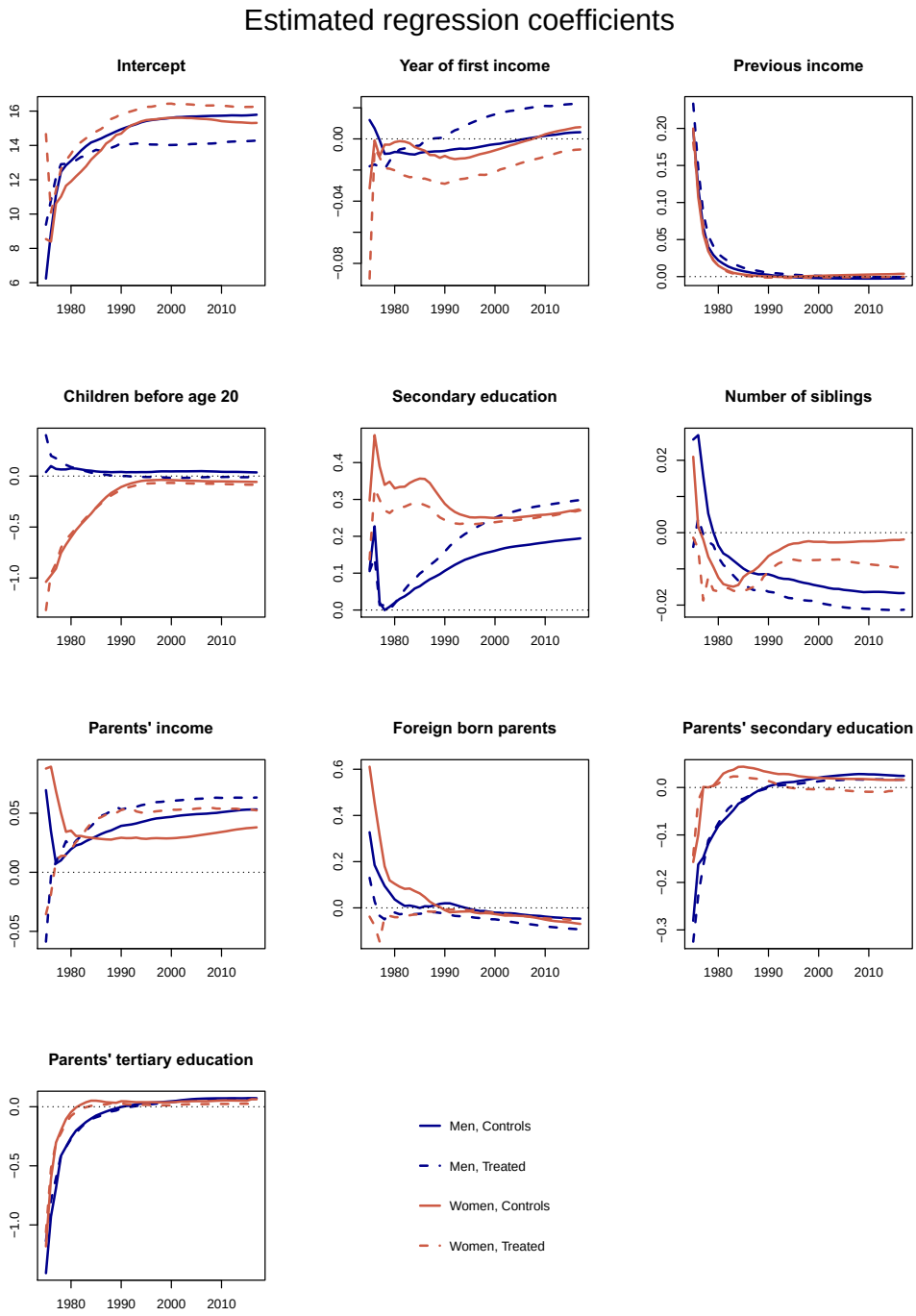

Figure E.1. Estimated regression coefficients for the two treatment groups, for men and women.

Table E.1. Estimated coefficients (γ) from the propensity score models (logistic regression)

Variable	Men	Women
Intercept	**−2.194**	**3.872**
Year of first income	**−0.052**	**−0.124**
Previous income	−0.001	**0.054**
Children before age 20	0.020	**−0.378**
Secondary education	**−0.097**	**0.109**
Number of siblings	**−0.056**	**−0.052**
Parents' income	**0.510**	**0.407**
Foreign-born parents	**0.568**	**0.541**
Parents' secondary education	**0.674**	**0.690**
Parents' tertiary education	**0.851**	**0.933**

Note. Coefficients in bold are significant at $\alpha = 0.05$.

E.2 Average estimates and confidence bands

Figure E.2 shows the averages of the estimated FATE and the four different types of confidence bands, taken over 10,000 replicates of the simulation study in Section 4. The results are for the male subset and a sample size of $n = 500$ in the fixed design setup, corresponding to the second column in Table 1.

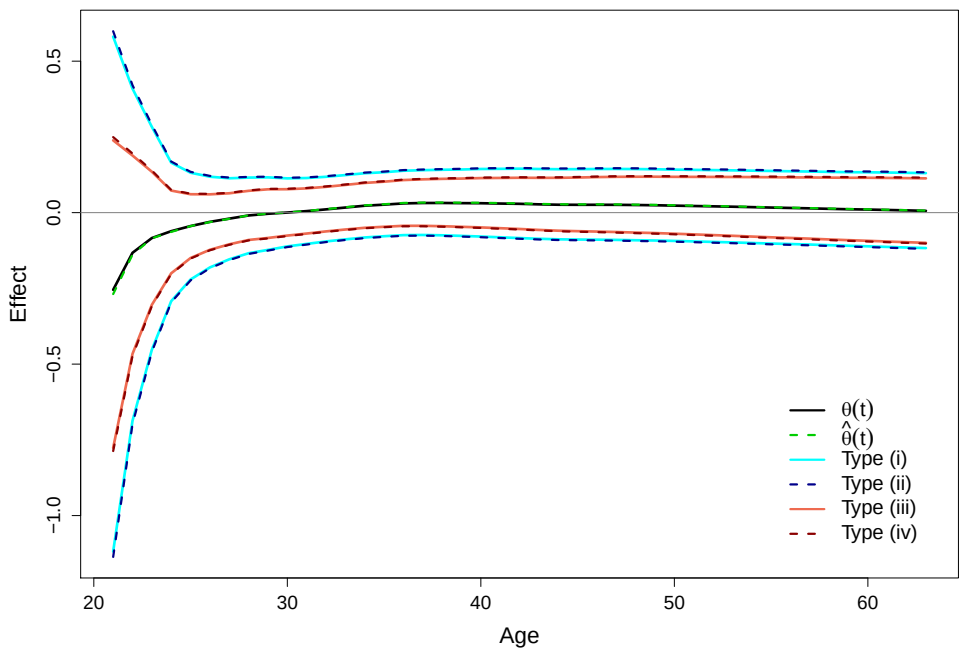

Figure E.2. Means of the FATE estimates and four types of confidence bands over the 1,000 simulation replicates for male subset, $n = 500$.

References

Statistics Sweden (2010). *Lokala arbetsmarknader – egenskaper, utveckling och funktion*. Statistics Sweden.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://R-project.org.

Abramowicz K., Häger C. K., Pini A., Schelin L., Sjöstedt de Luna S., & 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. https://doi.org/10.1111/sjos.v45.4

Altonji J. G., Kahn L. B., & 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. https://doi.org/10.1086/682938

Åslund O., & Rooth D. (2007). Do when and where matter? Initial labour market conditions and immigrant earnings. *The Economic Journal*, *117*(518), 422. https://doi.org/10.1111/j.1468-0297.2007.02024.x

Belloni A., Chernozhukov V., Fernández-Val I., & Hansen C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, *85*(1), 233–298. https://doi.org/10.3982/ECTA12723

Chen Z., Fan J., & Wang K. (2021). 'Remarks on multivariate Gaussian process', arxiv, arxiv:2010.09830, preprint: not peer reviewed.

Chen Z., Wang B., & Gorban A. N. (2020). Multivariate Gaussian and Student-*t* process regression for multi-output prediction. *Neural Computing and Applications*, *32*(8), 3005–3028. https://doi.org/10.1007/s00521-019-04687-8

Chib S., & Jacobi L. (2007). Modeling and calculating the effect of treatment at baseline from panel outcomes. *Journal of Econometrics*, *140*(2), 781–801. https://doi.org/10.1016/j.jeconom.2006.07.014

Choi H., & 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. https://doi.org/10.1111/rssb.12239

Genbäck M., & de Luna X. (2019). Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. *Biometrics*, *75*(2), 506–515. http://dx.doi.org/10.1111/biom.13001

Genz A., Bretz F., Miwa T., Mi X., Leisch F., Scheipl F., & Hothorn T. (2020). *mvtnorm: Multivariate normal and t distributions*. (R package Version 1.1-1) https://CRAN.R-project.org/package=mvtnorm.

Gupta A., & Nagar D. (2000). *Matrix variate distributions* (Vol. 104). Monographs and surveys in pure and applied mathematics. Chapman and Hall/CRC.

Holland P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. http://www.jstor.org/stable/2289064. https://doi.org/10.1080/01621459.1986.10478354

Holm S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Imbens G. W., & Rubin D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Jacobi L., Wagner H., & Frühwirth-Schnatter S. (2016). Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave. *Journal of Econometrics*, *193*(1), 234–250. https://doi.org/10.1016/j.jeconom.2016.01.005

Johansson P.-O., & Kriström B. (2016). *Cost-benefit analysis for project appraisal*. Cambridge University Press.

Karlsson C., & Olsson M. (2006). The identification of functional regions: Theory, methods, and applications. *The Annals of Regional Science*, *40*(1), 1–18. https://doi.org/10.1007/s00168-005-0019-5

Kwon I., Meyersson Milgrom E., & Hwang S. (2010). Cohort effects in promotions and wages evidence from Sweden and the United States. *Journal of Human Resources*, *45*(3), 772–808. https://doi.org/10.1353/jhr.2010.0017

Liebl D., & Reimherr M. (2019). *ffscb: fast n' fair simultaneous confidence bands for functional parameters*. (R package Version 0.0.10) https://github.com/lidom/ffscb.

Liebl D., & Reimherr M. (2020). Simultaneous inference for function-valued parameters: A fast and fair approach. In G. Aneiros, I. Horová, M. Hušková, & F. Vieu (Eds.), *Functional and High-Dimensional Statistics and Related Fields* (Chap. 21) Springer Nature Switzerland AG.

Liebl, D., & Reimherr, M. (2023). Fast and fair simultaneous confidence bands for functional parameters. Manuscript in preparation.

Lin Z., Kong D., & Wang L. (2021). 'Causal inference on distribution functions' arxiv, arxiv:2101.01599, preprint: not peer reviewed.

Lindgren U., Nilsson K., de Luna X., & 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), 1075–1075. https://doi.org/10.1093/ije/dyv358

Oreopoulos P., von Wachter T., & Heisz A. (2012). The short and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, *4*(1), 1–29. http://dx.doi.org/10.1257/app.4.1.1

Pini A., & Vantini S. (2017a). *fdatest: Interval testing procedure for functional data*. (R package Version 2.1.0) https://github.com/alessiapini/fdatest.

Pini A., & Vantini S. (2017b). Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, *29*, 407–424. https://doi.org/10.1080/10485252.2017.1306627

Raaum O., & 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. https://doi.org/10.1162/rest.88.2.193

Ramsay J. O., Graves S., & Hooker G. (2020). *fda: Functional data analysis*. (R package Version 5.1.5.1) https://CRAN.R-project.org/package=fda.

Ramsay J. O., & Silverman B. W. (2005). *Functional data analysis* (2nd ed.). Springer Series in Statistics. Springer.

Rosenbaum P., & Rubin D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Schwandt H., & 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. https://doi.org/10.1086/701046

Tan Z. (2007). Comment: Understanding or, PS and DR. *Statistical Science*, *22*(4), 560–568. https://doi.org/10.1214/07-STS227A

Telschow F., & Schwartzman A. (2022). Simultaneous confidence bands for functional data using the Gaussian kinematic formula. *Journal of Statistical Planning and Inference*, *216*, 70–94. https://doi.org/10.1016/j.jspi.2021.05.008

van der Laan M. J., & Rose S. (2018). *Targeted learning in data science: Causal inference for complex longitudinal studies*. Springer.

Vsevolozhskaya O., Greenwood M., & 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. https://doi.org/10.1214/14-AOAS723

Wang J.-L., Chiou J.-M., & Müller H.-G. (2016). Functional data analysis. *The Annual Review of Statistics and Its Application*, *3*(1), 257–295. https://doi.org/10.1146/statistics.2016.3.issue-1

Wooldridge J. F. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Massachusetts Institute of Technology.