ORIGINAL RESEARCH ARTICLE

Instrumental Variable Estimation for Functional Concurrent Regression Models

Justin Petrovich[a], Bahaeddine Taoufik[b], and Zachary Davis[c]

[a]Department of Business Administration, Saint Vincent College, Latrobe, PA, USA;
[b]Department of Mathematics, Saint Joseph's University, Philadelphia, PA, USA;
[c]Department of Economics, Saint Vincent College, Latrobe, PA, USA

ARTICLE HISTORY
Compiled February 21, 2023

ABSTRACT
In this work we propose a functional concurrent regression model to estimate labor supply elasticities over the years 1988 through 2014 using Current Population Survey data. Assuming, as is common, that individuals' wages are endogenous, we introduce instrumental variables in a two-stage least squares approach to estimate the desired labor supply elasticities. Furthermore, we tailor our estimation method to sparse functional data. Though recent work has incorporated instrumental variables into other functional regression models, to our knowledge this has not yet been done in the functional concurrent regression model, and most existing literature is not suited for sparse functional data. We show through simulations that this two-stage least squares approach greatly eliminates the bias introduced by a naive model (i.e., one that does not acknowledge endogeneity) and produces accurate coefficient estimates for moderate sample sizes.

KEYWORDS
functional concurrent regression; sparse functional data; instrumental variable; labor supply elasticity

1. Introduction

In this paper, we study labor supply elasticities using a functional concurrent regression model in which the functional data are observed sparsely. In economic theory, labor supply elasticities measure a person's response, in terms of hours worked, to a change in that person's hourly wage. Estimating labor supply elasticities for different subsets of the population provides insight into how different types of people value their leisure time over time spent working. Labor supply elasticities also, in part, determine how a change in income tax rates will affect production. To better understand questions like this, it is useful to study changes in labor supply elasticities over time.

Due to the centrality of labor supply elasticities in economic theory, there is an extensive literature studying them in various contexts. For some classic examples, see [2, 6, 24, 34]. More recently, Eissa [17], Eissa and Hoynes [18], and Aaronson and French [1] used changes in tax policies to estimate labor supply elasticities. Economists typically expect labor supply elasticities to be positive. Yet, using a novel dataset

CONTACT Justin Petrovich. Email: justin.petrovich@stvincent.edu

consisting of New York City cabdrivers, Camerer et. al. [11] showed that certain groups of workers have a negative labor supply elasticity. While the aforementioned papers all estimate labor supply elasticities at a particular point in time, Heim [25], Blau and Kahn [9], and Eckstein et al. [15] all estimate labor supply elasticities in different years using the same method across years.

Heim [25] showed that, over the 25 years from 1978 to 2002, the labor supply elasticity of married women shrunk significantly. He used a simultaneous equation model with four estimation stages to study the changes in hours worked in response to a change in wage. Other researchers, like MaCurdy [34], use structural estimation techniques. In this context, structural estimation introduces data into results from theoretical models to estimate parameters in the model for which data cannot be collected. Another method, introduced by Angrist [3], uses a grouped-data estimation technique to estimate labor supply elasticities that is similar to using an instrumental variable. More recently, Eckstein et al. [15] used a method of simulated moments to estimate their model life cycle decisions for both individuals and households in which it is necessary to estimate labor supply elasticities.

The type of longitudinal/panel data used by Heim and others can, however, fit into the framework of functional data analysis (FDA), a relatively recent sub-field of statistics [29, 38, 39]. The basic unit of data in FDA is a function and the simplest and most common examples are functions observed on a domain of time. Sample functions are typically assumed to come from a smooth population process, but are necessarily observed at discrete points on some grid, the number of which can range from one or a few (sparse) to tens or hundreds or more (dense). Thus, from an FDA perspective, longitudinal data is simply sparse functional data. Relationships among different functional variables are commonly studied via functional linear regression models [36]. Like other models traditionally applied to longitudinal data (e.g., mixed effects models), functional linear regression models and other FDA techniques are specifically designed to account for the inherent dependencies within functions, but typically assume smooth underlying processes or relationships.

Classical linear regression models (for scalar data) assume that the predictor variables are exogenous, i.e., that they are uncorrelated with the model's error term. However, for research in economics and other fields this is frequently an unreasonable assumption for a variety of reasons (e.g., measurement error in the predictor variable, omitted variables, etc.), leading to biased ordinary least squares estimators of the regression coefficients for endogenous variables (i.e., those which are correlated with the error term). Instrumental variables are commonly used to ameliorate this bias [4, 5, 26, 33]. The dearth of literature—particularly until the last several years—adapting instrumental variable (IV) estimation or other statistical methods to address the common problem of endogeneity for functional data represents a significant barrier to the adoption of FDA techniques by economists and other non-statisticians. Of those that have attempted to bridge this gap, most to date have focused on scalar-on-function regression models, where the response variable is scalar but some or all of the predictors are functional.

As an example, Florens and Van Bellegem [19] studied instrumental variable estimation in a scalar-on-function regression model with endogenous predictors. Using functional instrumental variables, a penalized least squares (PLS) method was applied to estimate the functional coefficients and convergence rates were established for it. Babii [7] showed that the same model can be identified using a real-valued (as opposed to a functional) instrumental variable. He established convergence rates for estimators based on both Tikhonov and Galerkin regularization and illustrated that the Galerkin estimator is minimax optimal. In another example, Tekwe et al. [44] developed an estimation strategy for a scalar-on-function model in which the functional predictor variable is contaminated by measurement error. To ameliorate the measurement error, a functional instrumental variable was used and a generalized method-of-moment-based estimator derived to estimate the functional coefficient. The estimator was shown to be L_2 consistent.

Studying a similar model setup to Tekwe et al. [44], Chen et al. [13] developed and established convergence rates for an autocovariance-based generalized method-of-moments (AGMM) estimator for the slope function using functional instrumental variables. They further extended this approach to the (fully) functional linear model (FLM), in which both the response and predictor are functional. Instrumental variables were also discussed in the context of the FLM by Benatia et al. [8] and Seong and Seo [43]. The former developed an estimation procedure for the bivariate regression coefficient using Tikhonov regularization while Seong and Seo proposed using a spectral cut regularized inverse of the covariance operators. The latter adapted this approach to both an IV estimator and a functional two-stage least square estimator.

Despite these recent developments, there are still some noticeable holes in the literature using functional regression models with endogenous predictors. First, while the aforementioned literature focused on scalar-on-function regression and the FLM, we focus instead on a functional concurrent regression (FCR) model. The FCR model involves both functional response and predictor variables, but is a constrained version of the FLM in that variables are assumed to be related contemporaneously.

Second, only one of the above-cited papers considered sparse functional data, for which each curve is observed at a small number of grid points. Because classical FDA techniques do not perform well when applied to sparse functional data, there has been considerable research over the past two decades into developing analytical methods appropriate for sparse functional data [12, 28, 30, 46]. Chen et al. [13] did investigate both the asymptotic and finite-sample (through simulations) properties of the proposed instrumental variable estimator when the functional predictors are partially observed, but this was done only in a scalar-on-function regression model. While there is a growing literature addressing FCR models for sparse functional data [21, 22, 31, 42, 47], to our knowledge no work to date has allowed for endogenous predictors.

Motivated by the problem of estimating changing labor supply elasticities over time, we add to the nascent literature on instrumental variable estimation in functional regression models by filling both of the aforementioned holes. Specifically, we propose an algorithm for estimating the functional coefficient of an endogenous functional predictor variable in an FCR model. Importantly, unlike current methods, this algorithm is designed and performs well for sparse functional data. The usefulness of this estimation procedure is demonstrated by estimating labor supply elasticity curves, which simultaneously highlights the utility of applying functional data methods in studying such problems.

The rest of this paper is organized as follows. The Current Population Survey data, which are used to estimate the labor supply elasticities of interest, are introduced in Section 2. Section 3 presents the proposed model and notation, followed in Section 4 by an outline of our estimation procedure and an algorithm for implementing it. In Section 5, the performance of the proposed estimation procedure is studied through simulations. In Section 6, we return to the problem of estimating labor supply elasticities and apply our proposed estimation procedure. Finally, concluding thoughts and ideas for future study are offered in Section 7.

2. Current Population Survey Data

To estimate labor supply elasticities, we use data from the Current Population Survey (CPS), years 1988–2014. The CPS is administered by the U.S. Census Bureau for the Bureau of Labor Statistics and surveys approximately 60,000 U.S. households each month. The primary purpose of the CPS is to measure the size of the labor force and the status of the labor force participants. The data are used to calculate the unemployment rate each month. While the CPS contains basic employment status and demographic data, the survey is regularly supplemented with questions on other topics. The March supplement includes questions about income received in the previous year, which is of particular interest.

The CPS surveys the same respondent for four consecutive months, gives the respondent an eight month break, then returns to survey the respondent for four more months. Since the March supplement of the CPS is the only month that contains the detailed income data we need to estimate labor supply elasticities, the method used by the CPS to collect the data yields two data points for each person in the data set. While on the one hand, this design is limiting due to the paucity of observations it provides for each respondent, the large number of households surveyed does guarantee a large number of observations in any one year. Classical functional data methods do not perform well with such designs, but sparse functional data methods leverage the entire sample size and pool information from across subjects to overcome this difficulty. What is important for sparse functional data methods is a dense sampling over the observed grid for the functions when all observations for all respondents are combined.

In our analysis, we consider only working adults, aged 25 to 65 years old. After removing respondents who don't fit this criterion, as well as those missing data for the model variables (see Section 6), our data include 245,531 respondents. The distribution of respondents per year is shown in Figure 1. Apart from the years 1992–1995, in no year are there fewer than 3,000 respondents. The years 1992 through 1995, however, include only 61, 7, 33, and 28 respondents, respectively. The low number of respondents in these four years is due to missing data that is necessary to calculate the marginal tax rates in the CPS. We calculate the marginal tax rate for each individual using TAXSIM9, a program created by the National Bureau of Economic Research. These tax rates are then used to calculate the after tax hourly wage for each individual. While the estimated LSEs for 1992–1995 will carry more uncertainty than the others, with thousands of people sampled in most years, the deficiency of observing each person only twice can be overcome by applying sparse functional data methods and pooling across respondents.

3. Model and Notation

FCR models consist of a functional response variable and at least one functional predictor, measured on the same domain (often, but not limited to time), as well as the possibility of other scalar (i.e., non-functional) covariates. The FCR model can be written as follows:

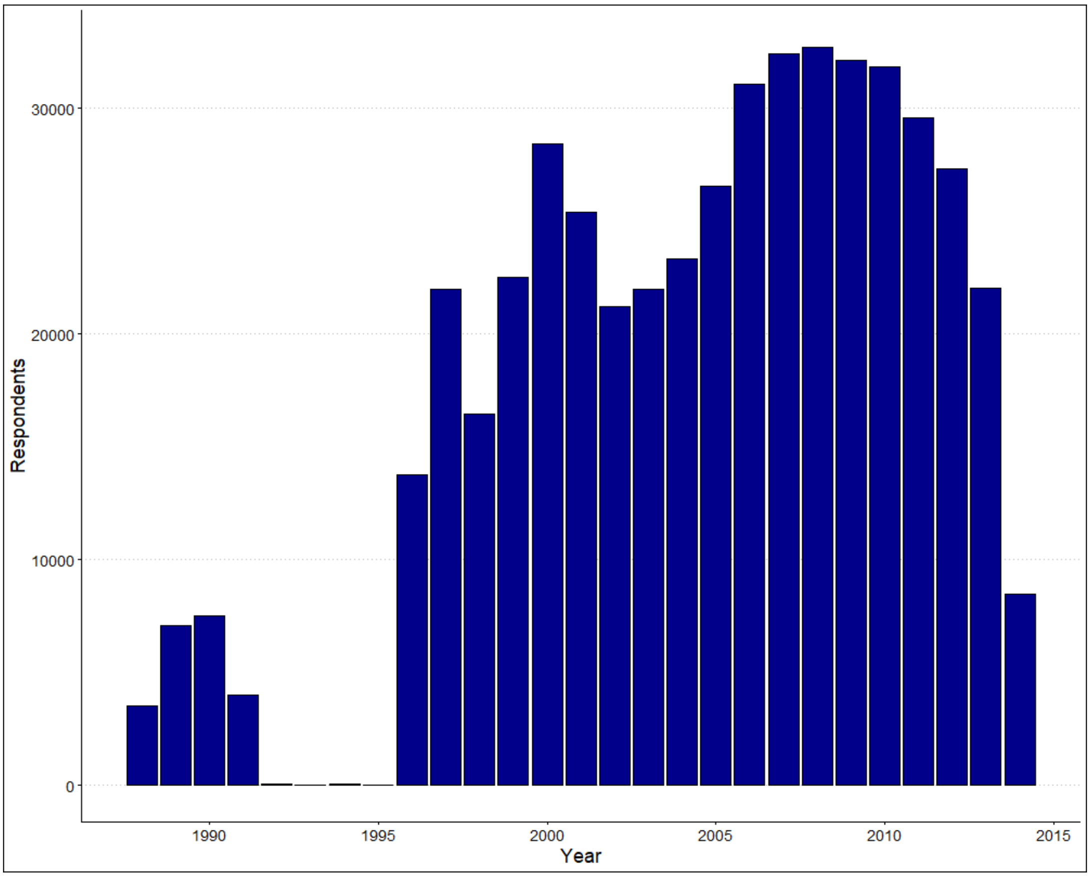

Figure 1. Number of CPS respondents per year included in the data after all inclusion criteria were applied.

$$Y(t) = \beta_0(t) + \sum_{p=1}^{K_1} \beta_p(t) X_p(t) + \alpha(t) W(t) + \sum_{q=1}^{K_2} \theta_q S_q + \varepsilon(t). \tag{1}$$

The response variable is denoted by $Y(t)$ and the $\{X_p(t), W(t) : p = 1, \ldots, K_1\}$ are functional covariates, all assumed to be square-integrable functions belonging to the space $L^2[0, 1]$. In this setup, t represents the argument to the function, e.g., time. For simplicity, we assume that $t \in [0, 1]$, although the functional domain need not be restricted as such in general. The random variables $\{S_q : q = 1, \ldots, K_2\}$ represent scalar covariates in the model and $\varepsilon(t) \in L^2[0, 1]$ is the random error term. The function $\beta_0(t)$ is the functional regression intercept and $\beta_1(t), \ldots, \beta_{K_1}(t)$, and $\alpha(t)$ are all unknown functional regression coefficients. The parameters $\{\theta_q : q = 1, \ldots, K_2\}$ represent regression coefficients of the scalar variables.

A distinguishing feature of Model (1) from past literature on FCR models is that the functional predictor variable, $W(t)$, is assumed to be endogenous. The other functional covariates, $X_p(t), p = 1, \ldots, K_1$, are assumed to be exogenous, as is typical. Specifically, we assume that $Cov[X_p(t), \varepsilon(s)] = 0$ while $Cov[W(t), \varepsilon(s)] \neq 0$, for all p and $s, t \in [0, 1]$. For simplicity, we also assume that the scalar covariates are exogenous, i.e., $Cov[S_q, \varepsilon(t)] = 0$ for all q and all $t \in [0, 1]$. Since methods for consistently estimating the functional coefficients of the exogenous variables have been well established, our focus is on appropriately estimating the coefficient for the endogenous functional predictor, i.e., $\alpha(t)$.

In practice, we observe N sample realizations of the model variables. For the functional variables, we assume these are random draws from the underlying population processes, making them independent realizations of smooth, random functions. However, we additionally assume that the functional data are observed sparsely over the domain. This is typically taken to mean that the functions are observed with measurement error at a relatively small number of randomly-sampled locations along the domain. Thus, the observed values of the i^{th} subject at time j can be written as

$$\begin{aligned} y_{ij} &= Y_i(t_{ij}) + \delta_{ij} \\ x_{ijp} &= X_{ip}(t_{ij}) + e_{ijp} \\ w_{ij} &= W_i(t_{ij}) + \nu_{ij} \end{aligned}$$

all for $1 \leq i \leq N$ and $1 \leq j \leq m_i$, where m_i denotes the number of observations for the i^{th} subject. The terms δ_{ij}, e_{ijp}, and ν_{ij} represent independent, identically distributed measurement errors which all have mean 0 and finite variances, $Var(\delta_{ij}) = \sigma_\delta^2$, $Var(e_{ijp}) = \sigma_{e_p}^2$, and $Var(\nu_{ij}) = \sigma_\nu^2$. We further assume that both the number and the time points of observations for each subject, m_i and t_{ij}, respectively, are random (e.g., uniformly distributed). Though the delineation between sparsely and densely sampled functional data is somewhat subjective, here we take sparse to mean that most subjects are observed only a few times (e.g., less than 10). We additionally assume that m_i, T_{ij}, and the measurement errors, δ_{ij}, e_{ijp}, and ν_{ij} are all mutually independent, as well as independent of the random functions $\{Y(t), X_p(t), W(t) : p = 1, \ldots, K_1\}$. We begin with these classic sparse functional data assumptions but explore some deviations in the simulations of Section 5 to better match our data application in Section 6.

4. Estimation Method

In this section, we propose an estimation approach analogous to the Two-Stage Least Squares (2SLS) estimator commonly used in scalar regression models [23]. The idea hinges on the existence of one or multiple instrumental variables, variables that are correlated with the endogenous variable but which are themselves exogenous. That is, we assume there exist random functions $\{Z_r(t) \in L^2[0,1] : r = 1, \ldots, K_3\}$ such that $Cov[W(s), Z_r(t)] \neq 0$ and $Cov[Z_r(t), \varepsilon(s)] = 0$ for all r and $s, t \in [0,1]$. These assumptions guarantee the relevance and validity, respectively, of the instruments. Our two-stage least squares estimation strategy can then be summarized by the following algorithm.

(1) Impute the sparsely observed functional response variable, $Y(t)$ for each observation, $i = 1, \ldots, N$. We do this by using PACE, the approach developed by Yao et al. [46]. PACE is used for its simplicity and applicability—it works well in the sparsest situations, even when some curves may be observed at only a single time. First, define $\mu_Y(t) = E[Y(t)]$ and $C_Y(s,t) = E[(Y(t) - \mu_Y(t))(Y(s) - \mu_Y(s))]$, the mean and covariance functions of $Y(t)$. PACE leverages the Kharunen-Loève

expansion, which can be written as follows:

$$Y_i(t) = \mu_Y(t) + \sum_{p=1}^{\infty} v_p(t)\xi_{ip}. \tag{2}$$

In this expression, the $v_p(t)$ are the eigenfunctions of C_Y, with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, and the scores are defined as $\xi_{ip} = \langle Y_i - \mu_Y, v_p \rangle$. Truncating the sum and estimating each of the quantities on the left-hand side of Equation (2), an approximation can be found for each $Y_i(t)$ by using the following steps.

(a) First, obtain an estimate, $\hat{\mu}_Y(t)$, of the mean function of Y. This can be obtained a number of different ways, but PACE uses local linear scatterplot smoothing to obtain a smooth estimate.

(b) Next, obtain an estimate, $\hat{C}_Y(s,t)$ of the covariance function of Y. This can again be obtained a number of different ways, but we follow PACE and obtain a smooth estimate by applying a two-dimensional scatterplot smoother.

(c) Perform functional principal component analysis (FPCA) on $\hat{C}_Y(s,t)$, and obtain estimates $\hat{v}_p$ and $\hat{\lambda}_p$ of the eigenfunctions and eigenvalues, respectively, for $p = 1, \ldots, P$. The number, P, is chosen by selecting the minimum number of principal components that explain at least 95% of the variance in $Y(t)$.

(d) Use PACE to obtain the best linear unbiased predictors of the ξ_{ip}. Call these $\hat{\xi}_{ip}$ for $p = 1, \ldots, P$.

(e) Finally, an approximation of $Y_i(t)$ can be found by plugging in the estimators to Equation (2) and truncating the sum at P:

$$\tilde{Y}_i(t) \approx \hat{\mu}_Y(t) + \sum_{p=1}^{P} \hat{v}_p(t)\hat{\xi}_{ip}.$$

(2) Follow Step 1 for each functional predictor variable. Call these imputed functions $\tilde{X}_1(t), \ldots, \tilde{X}_{K_1}(t), \tilde{W}(t)$.

(3) Follow Step 1 for the instrumental variables and denote these imputed functions by $\tilde{Z}_1(t), \ldots, \tilde{Z}_{K_3}(t)$. At conclusion, you will have imputed estimates of all functional variables observed as densely as desired.

(4) Use Two-Stage Least Squares to obtain an estimate of $\alpha(t)$, denoted as $\hat{\alpha}(t)^{2SLS}$, along with estimates of the other regression coefficients. At each stage, the coefficient functions are estimated using the penalized functional regression approach introduced by Ivanescu et al. [27] and implemented in the R package, *refund*. These steps can be outlined as follows:

(a) Stage 1: regress $\tilde{W}(t)$ on all remaining predictor variables as well as the instrumental variables. That is, estimate the model

$$\tilde{W}(t) = \gamma_0(t) + \sum_{p=1}^{K_1} \gamma_p(t)\tilde{X}_p(t) + \sum_{q=1}^{K_2} \eta_q S_q + \sum_{r=1}^{K_3} \phi_r(t)\tilde{Z}_r(t) + \nu(t) \tag{3}$$

From this regression, obtain the predicted curve, $\hat{W}(t)$.

(b) Stage 2: Estimate Model (1), but using the imputed functional variables and replacing $W(t)$ with $\hat{W}(t)$. That is, estimate the model

$$\tilde{Y}(t) = \beta_0(t) + \sum_{p=1}^{K_1} \beta_p(t)\tilde{X}_p(t) + \alpha(t)\hat{W}(t) + \sum_{q=1}^{K_2} \theta_q S_q + \varepsilon(t) \tag{4}$$

The same estimation process is used as in Stage 1. The estimate for $\alpha(t)$ from this model is $\hat{\alpha}(t)^{2SLS}$.

Models (3) and (4) are estimated in much the same way. For instance, for Stage 1, begin by applying a basis expansion to the coefficients, $\gamma_0(t), \gamma_1(t), \ldots, \gamma_{K_1}(t), \phi_1(t), \ldots, \phi_{K_3}(t)$. Ivanescu et al. [27] advocate using a large number of basis functions and then smoothing the coefficient functions by penalizing them. The default settings in the *refund* packages implementation (i.e. the pffr function) employ cubic b-splines and a first order difference penalty, as described by Eilers and Marx [16], but many options are possible for the choice of basis and penalization. For more details about the penalization of Models (3) and (4), see Appendix, Section A.

5. Simulation Study

5.1. Simulation Set-up

In this section, Monte Carlo simulations are used to evaluate the performance of the FC2SLS model and to compare it against a naive FCR model that ignores the endogeneity issue. In our simulated model, we suppose that the response variable is determined by two predictor variables, one of which is unobserved and, thus, not included in estimation. So long as the predictor variables are correlated, omitting one is a common source of endogeneity in the remaining predictor, resulting in both biased and inconsistent estimates of the regression coefficient. In these simulations, we explore the same effects in a functional concurrent regression model.

Simulations are based on Model 1, where $K_1 = 1$ and $K_2 = 0$ for simplicity. Written explicitly:

$$Y(t) = \beta_0(t) + \beta_1(t)X_1(t) + \alpha(t)W(t) + \varepsilon(t)$$

where the endogenous functional predictor, W, is simulated as

$$W(t) = \delta_0(t) + \delta_1(t)W^*(t) + \delta_2(t)Z(t)$$

and

$$W^*(t) = \theta_0(t) + \theta_1(t)X_1(t) + \lambda(t).$$

The error terms are assumed to be independent such that $\varepsilon(t) \sim N(0, \sigma_\varepsilon^2)$ and $\lambda(t) \sim N(0, \sigma_\lambda^2)$ for all $t \in [0, 1]$. Throughout all simulations, $\sigma_\varepsilon = 1$ and $\sigma_\lambda = 0.5$ are used.

The random function X_1, which is unobserved and omitted from the fitted model (causing W to be endogenous), is simulated as a Gaussian process with mean function $\mu_{X_1}(t) \equiv 0$ for all time points and covariance function defined by the following Matérn covariance function

$$C_{X_1}(t,s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}|t-s|}{\rho} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}|t-s|}{\rho} \right),$$

where K_ν is the modified Bessel function of the second kind and the parameters are chosen as $\nu = 5/2$, $\rho = 0.5$, and $\sigma^2 = 10$. The variable Z is the instrumental variable, which is assumed to be pointwise uncorrelated with ε, λ, and X_1 and is also simulated as a mean-zero Gaussian process with covariance function $C_Z(s,t)$.

Across all simulations, functions are generated over a grid of $M = 100$ evenly-spaced time points from [0,1]. In each scenario, we use two different sampling schemes, the first following the classical sparse functional data assumptions and the second tailored to match the CPS data that we study in Section 6. These sampling designs can be summarized as follows.

- **Classic sparse design (CSD)**: For each curve, $i = 1, \cdots, N$, the number of observations per curve, m_i, is a random variable drawn from a discrete uniform distribution $U[2,5]$. In addition, the observed grid points themselves, $t_{i_1}, \cdots, t_{i_{m_i}}$ are sampled uniformly from the grid of 100 evenly-spaced time points in [0,1].
- **CPS design (CPS)**: Each curve receives either 1 or 2 observations according to the following rules. For each curve, $i = 1, \cdots, N$, an initial observation is randomly selected from the evenly-spaced grid of time points on [0,1], call it $t_{i_1} = t^*$. If $t^* = 1$, then the i^{th} curve does not receive a second observation. Otherwise, the i^{th} curve receives a second observation and $t_{i_2} = t^* + \frac{1}{99}$, where $\frac{1}{99}$ is the time difference between any two points on the evenly-spaced grid.

Three different model scenarios, described in the following subsections, are studied. Each scenario is simulated 100 times for multiple sample sizes (N = 100, 200, 400, and 800). For each simulation, we focus on estimating $\alpha(t)$ and compare estimates from four different methods. A naive estimate uses W directly to predict Y without correcting for the endogeneity of W. Functional concurrent regression was applied to obtain naive estimates of α using both the *fcr* [32] and the *fdapace* [20] R packages. The former allows for a functional random error term and uses a spline expansion and generalized penalized least squares to estimate the regression coefficients. The latter is based on the local linear smoothing approach employed by PACE and also outlined by Şentürk and Nguyen [42]. We label these two approaches *Naive-f* and *Naive-p*.

FC2SLS estimation is carried out as described in Section 4, but using two different methods to perform the imputation described in the first three steps. The first of these, which we call *FC2SLS-p*, again uses the *fdapace* package, which implements the PACE procedure outlined in Section 4. The second, which we call *FC2SLS-f* is based on the FACE algorithm derived by Xiao et al. [45], which estimates the covariance function and predicts (i.e. imputes) curves using a P-spline approach. Aside from the differences in imputation, *FC2SLS-p* and *FC2SLS-f* proceed identically and use the *refund* package to estimate coefficients in both stages of regression.

Performance of each estimation method is measured according to the Median Integrated Squared Error (ISE) of its estimated coefficient function, defined as $median \int (\hat{\alpha}_s(t) - \alpha(t))^2 dt$, where the median is taken over simulations $s = 1, \ldots, 100$. Mean ISEs were also calculated and are reported for reference in the Appendix, Section B.

5.2. *Scenario 1: Constant Functions, One Instrument*

The first model highlights the connection to scalar regression models by assuming that all functions are constant over time. In this case, we expect naive and two-stage least square estimators of regression coefficients to perform similarly to how they would in a simple scalar regression model.

The regression coefficients are simulated as follows: for all time points in the evenly-spaced grid of 100 points from 0 to 1, $\beta_0(t) \equiv 1$, $\beta_1(t) \equiv 1$, and $\alpha(t) \equiv 1$. In addition, W is generated with coefficients $\delta_0(t) \equiv 0$, $\delta_1(t) \equiv 0.5$, and $\delta_2(t) \equiv 1$, and W^* with $\theta_0(t) \equiv 1$, $\theta_1(t) \equiv 0.5$. Finally, the covariance matrix of the instrumental variable, Z, is defined as a Rational Quadratic covariance function:

$$C_Z(t,s) = 0.5\left(1 + \frac{|t-s|^2}{2V\kappa^2}\right)^{-V},$$

where $\kappa = 1$ and $V = 2$.

Results are shown in Table 1, which reports the Median ISE. Clearly the FC2SLS estimation methods outperform the naive methods. Regardless of whether using the CSD or CPS sampling design, both FC2SLS methods have lower Median ISE than the naive methods for all sample sizes. Furthermore, the Median ISE for both FC2SLS methods approaches 0 as N grows from 100 to 800, suggesting consistency. This is unsurprising since the 2SLS estimator is consistent, while the naive estimator has asymptotic bias, in scalar regression models. One important caveat is that *FC2SLS-f* occasionally produced extraordinarily errant estimates, particularly for sample sizes of 100 and 200. This is actually a theme of the *FC2SLS-f* method and can be better seen reflected in the Mean ISEs reported in the Appendix, Section B.

5.3. *Scenario 2: Non-Constant Functions, One Instrument*

Unlike in the first simulation scenario, in this scenario we chose regression coefficients that varied over time. Specifically, the coefficients that define Y were chosen as $\beta_0(t) \equiv 1$, $\beta_1(t) = 1 + sin(2\pi t)$, and $\alpha(t) = 1 + cos(2\pi t)$.The remaining coefficient functions were defined identically as in Scenario 1. Referring again to Table 1, we see a similar story as in Scenario 1. For both sampling designs and all sample sizes, the two 2SLS estimators have lower Median ISEs than do the naive estimators. While the Median ISE for *FC2SLS-f* and *FC2SLS-p* stops improving when increasing N from 400 to 800, they still reach relatively low error levels, certainly much lower than both naive approaches. Since these Median ISE values aggregate 100 simulations in each setting, it is interesting to see an example estimate in each case as well.

Figure 2 shows three example simulations for each sample size under the CSD design in Scenario 2. These example plots illustrate the ability of *FC2SLS-f* and *FC2SLS-p* to capture the shape of the true coefficient function, as well as the bias that exists in both naive estimators. The first simulation shown for $N = 100$ highlights that for the smaller sample sizes used, wild estimates occasionally occurred, particularly for the *FC2SLS-f* method. However, these estimates were anomalous and vanished for the larger sample sizes of 400 and 800. It is also interesting to note that both naive methods in this scenario produced extremely similar—and biased—coefficient function estimates, regardless of the sample size.

Table 1. **Median** integrated squared error of $\hat{\alpha}(t)$ from each method across 100 simulations.

		CSD				CPS			
Scenario	N	Naive-f	Naive-p	FC2SLS-f	FC2SLS-p	Naive-f	Naive-p	FC2SLS-f	FC2SLS-p
1	100	1.02	4.05	0.12	0.10	1.06	4.16	0.21	0.12
	200	1.01	3.91	0.06	0.07	1.04	3.97	0.08	0.07
	400	1.00	3.96	0.03	0.04	1.01	4.07	0.04	0.05
	800	1.00	3.92	0.02	0.02	1.01	4.05	0.02	0.03
2	100	1.67	6.53	0.45	0.41	1.62	7.00	0.69	0.44
	200	1.66	6.40	0.38	0.36	1.66	6.56	0.69	0.34
	400	1.64	6.51	0.30	0.29	1.51	6.45	0.56	0.31
	800	1.65	6.51	0.31	0.29	1.53	6.60	0.58	0.30
3	100	1.28	1.34	0.21	0.16	1.67	1.25	0.45	0.28
	200	1.25	1.29	0.13	0.09	1.69	1.14	0.26	0.17
	400	1.13	1.15	0.11	0.08	1.63	1.06	0.21	0.14
	800	1.13	1.13	0.08	0.05	1.61	0.93	0.15	0.12

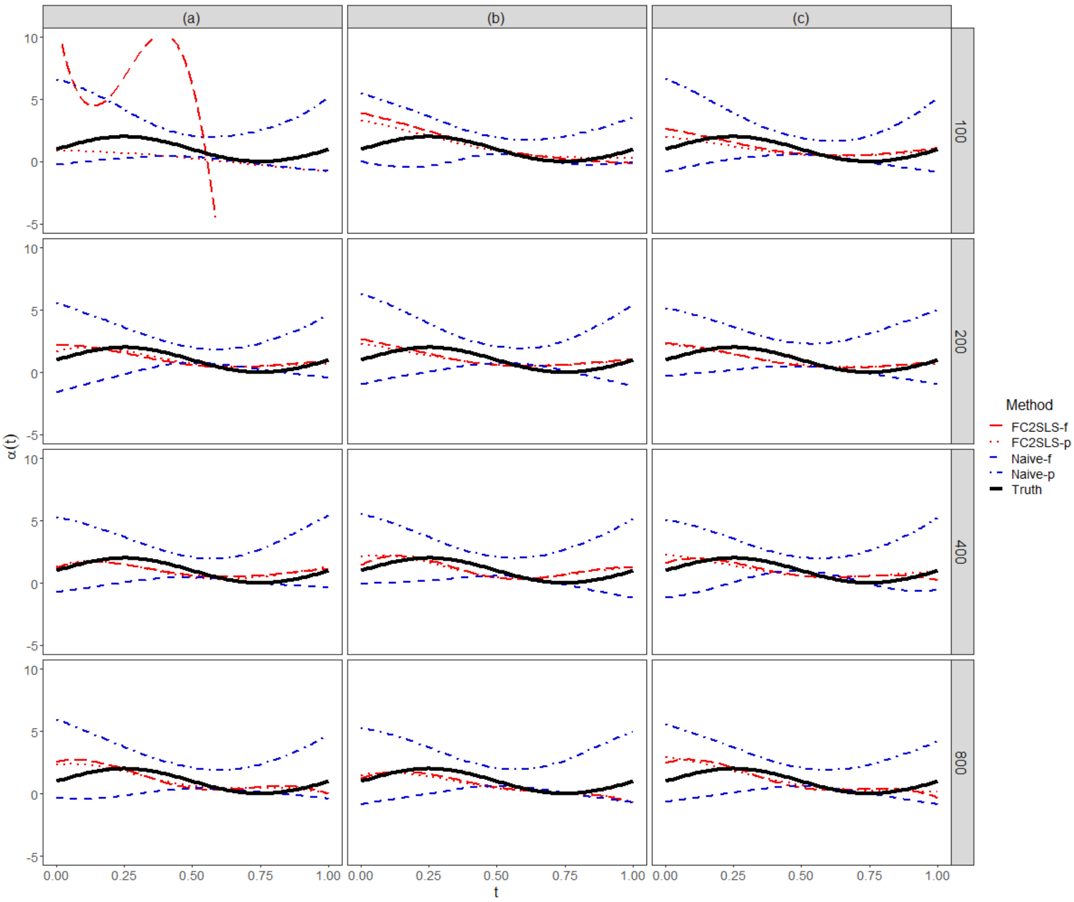

Figure 2. Estimated coefficient functions from three example simulations following Scenario 2 and the CSD sample scheme. Each row represents a different sample size ($N = 100, 200, 400,$ or 800) and each column represents one of the three example simulations (a, b, and c). Examples shown are simply the first three simulations for each scenario.

5.4. Scenario 3: Non-Constant Functions, Two Instruments

In the third and final simulation scenario, we additionally allowed the coefficients that define W and W^* to vary over time and made two other changes. First, the covariance of the instrument was defined to have the same Matérn covariance function as X_1 so that $C_Z(t, s) = C_{X_1}(t, s)$, this time with $\sigma^2 = 1$. Second, an additional instrumental variable, this one defined as $Z^2(t)$, was added to the model. The model for generating W was thus changed to

$$W(t) = \delta_0(t) + \delta_1(t)W^*(t) + \delta_2(t)Z(t) + \delta_3(t)Z^2(t).$$

The coefficients in the regression of Y were defined $\beta_0(t) \equiv 1$, $\beta_1(t) = 1+t\sin(2\pi t)$, and $\alpha(t) = 5+5t\cos(2\pi t)$. Next, we let $\delta_0(t) \equiv 0$, $\delta_1(t) = 1+0.5\sin(2\pi t)$, $\delta_2(t) = 1+\sin(2\pi t)$, and $\delta_3(t) = 1 + \cos(2\pi t)$. The coefficients for W^* were defined as $\theta_0(t) = 1 + \sin(2\pi t)$, $\theta_1(t) = 1 + 0.5\cos(2\pi t)$. This scenario is more realistic than the others and enables us to explore the benefit of using two instruments in the two-stage estimation approach.

From the results in Table 1, it is apparent that the 2SLS methods again dominate the naive methods. The Median ISE of both 2SLS methods also tends towards 0 again as the sample size grows. Figure 3 illustrates the estimated coefficient functions from all four methods in three of the simulations for Scenario 3 under the CSD sampling design. As in Scenario 2, *FC2SLS-f* occasionally produced extreme estimates for the smallest sample sizes. One example of this can be seen in the top left panel of Figure 3. On the other end of the sample size spectrum, it is clear from the three simulations shown with $N = 800$ that both 2SLS estimators were capable of capturing the true shape of the coefficient function. The naive methods, in contrast, occasionally capture the correct shape but again consistently demonstrate bias in their estimates.

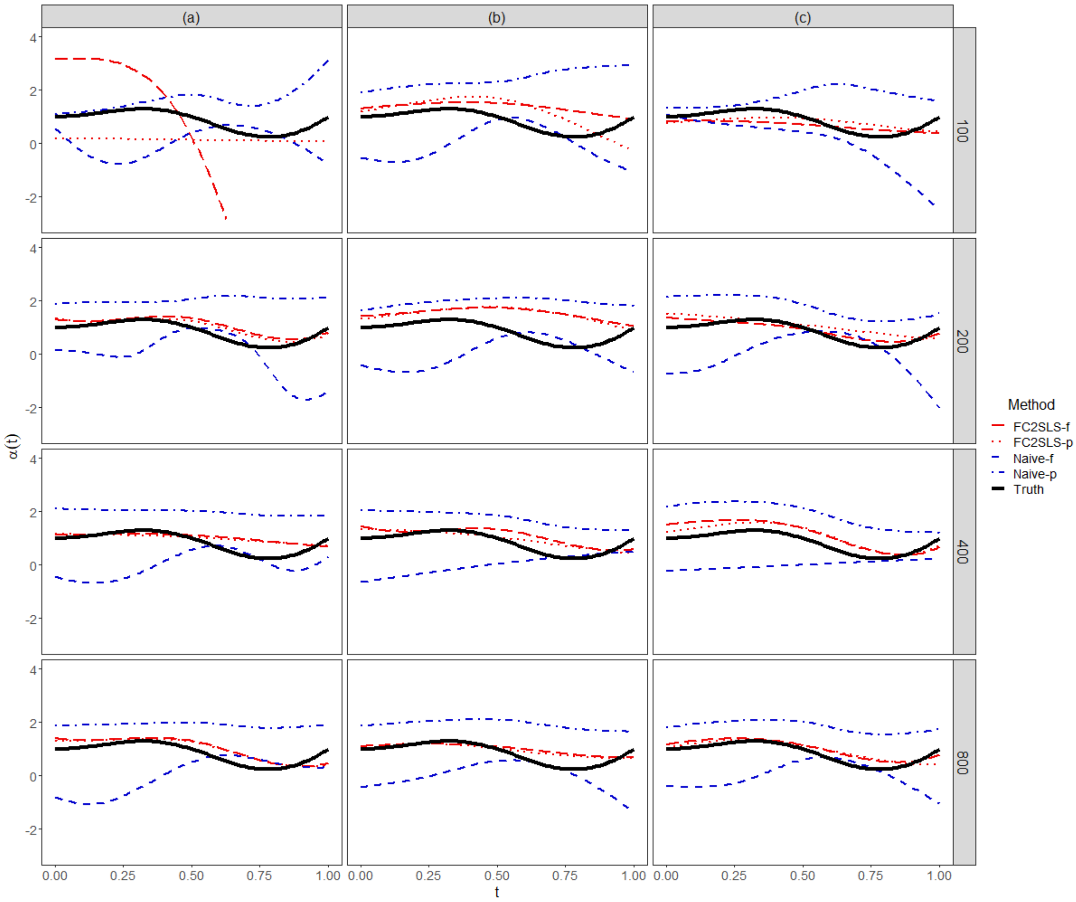

Figure 3. Estimated coefficient functions from three example simulations following Scenario 3 and the CSD sample scheme. Each row represents a different sample size ($N = 100, 200, 400,$ or 800) and each column represents one of the three example simulations (a, b, and c). Examples shown are simply the first three simulations for each scenario.

6. Analysis of Current Population Survey Data

Model 1 was applied to the CPS data, described in Section 2, using the proposed 2SLS estimation procedure. Note that the use of an FCR model in this case implies that the effects of a change in wages in year t correspond with a particular change in the amount of hours worked *in the same year*. This seems plausible given that a year is a large unit of time to accommodate shifting work behavior. However, to investigate the possibility that the a change in one's wages at time t may have a lagged impact on their hours worked, some formation of historical functional linear regression model could alternatively be used [35, 41]. We leave this exploration for future work and describe only the results for the FCR model.

In the notation of Model 1, $Y(t)$ represents the natural log of the number of hours worked per week in year t and $W(t)$ represents the natural log of after-tax wages in year t. Thus, $\alpha(t)$ can be interpreted as the labor supply elasticity at time t. Note, however, that wage is likely endogenous to hours. A person's preference to work may jointly determine the number of hours worked as well as their wage. Since preference for work is unobserved, the omission of this variable biases the elasticity estimate. In addition, the wage variable for salaried employees is calculated by taking their salary and dividing by hours worked. Calculating an independent variable by dividing by the dependent variable creates division bias in the estimate, as discussed in Borjas [10]. Hence an instrument is used for after-tax wages. In the spirit of [25], we use the quadratic function of age as the instrument in this case. This choice of instrument assumes that wages increase earlier in life but flatten when we get older. Hours, on the other hand, are assumed to have only a linear relationship with age. As Heim notes, instruments based on restrictions on how age and education affect hours worked have been criticized as likely being invalid. Using variation in tax rates are more plausible, but tax rates change infrequently. Our data spans many years, so we need an instrument that is consistent across those years. The quadratic age term satisfies that requirement and has been used previously as an instrument for wage.

Age and non-wage earnings were included as functional controls in the model, so $K_1 = 2$. In addition, occupation, education level, and number of kids were included in the model as linear (i.e., non-functional) controls, making $K_2 = 3$. Occupation was divided into 16 categories according to our own aggregation of a time consistent set of occupational categories developed by Dorn [14]. Dummy variables for each of the 16 occupational categories were included as controls. The CPS also divides education level into 17 categories, dummy variables for which were used as controls. For all individuals in the sample data, education level, occupation, and number of kids were constant across both observations, so these variables were all treated as scalar covariates (i.e., they had time-invariant effects on hours worked).

The model was fit for four demographic groups: married males, unmarried males, married females, and unmarried females. Due to the size of the data each group was randomly sampled if over 50,000 individuals were in the data set, so each sub-sample contained either all members of the subgroup or a sample of size 50,000, whichever was smaller. This resulted in sample sizes of 32,754 unmarried males; 50,000 married males; 40,125 unmarried females; and 50,000 married females. Estimated coefficient functions from each model were obtained using the FC2SLS estimation method outlined in Section 4.

Of particular interest are the estimated labor supply elasticities (LSE) as a function of year for each demographic subgroup, which are shown in Figure 4. The estimates reveal that the LSE of married males decreased sharply beginning in the early 2000s, transitioning from a value around 0.2 throughout the late 1980s to a peak of about 0.7 in 2003 and dropping to a low of nearly -0.6 by 2010 before rebounding sharply to 0.2 again by 2014. Unmarried males, on the other hand, saw much less change in their LSE, which slowly and steadily rose from 0 to about 0.3 in the early 2000's before dropping again to nearly 0 over the next decade. The LSE estimates for married females are far more striking and saw the most change out of the four groups. Not only did their LSE transition from negative to positive to negative again, but the magnitudes of these changes were quite large, especially the steep decline from about 1.2 in 2003 to -1.5 in 2013. Finally, unmarried females were estimated to follow a pattern similar to that of married males. For instance, their estimated LSE more than tripled from 0.4 in 1988 to a peak of over 1.3 in 2000 and 2001, after which it plummeted to a value of -0.3 by 2011 before returning nearly to 0 in 2014. Note that the conspicuous lack of confidence bands for all of these estimates is because the standard errors produced from the second stage model would neglecting two significant sources of uncertainty: the uncertainty from imputing the sparsely-observed functional variables and the uncertainty introduced by using the predicted wages in the second stage model.

Economic theory is silent on whether labor supply elasticities are positive or negative. When the hourly wage increases, leisure time becomes relatively more expensive causing the household to supply more labor. The same increase in the hourly wage also makes the household feel wealthier causing the household to supply less labor in order to enjoy the newly acquired wealth. Since these effects of an increased hourly wage occur simultaneously, the higher wage has an ambiguous effect on labor supply. Empirical evidence supporting both positive and negative labor supply elasticities exists [see, for example 11, 25]. Evidence from [25] and [15] also suggests married female labor supply elasticities have decreased over time, which is similar to our results.

Labor supply theories in economics generally have households choose the hours of labor to work in response to the wage. Wages are not influenced or chosen by households. Our model's specification with hours worked as the response variable and the wage as its predictor follows from this theory. In reality, employers may offer higher wages to people who are willing to work more hours. Though similar to the unobserved preference for work, this potential reverse causality of hours on wages is another possible source of endogeneity in our model. Our instrumental variables method corrects for the bias introduced by reverse causation.

The sign and magnitude of the labor supply elasticities for these different groups helps predict how these groups will respond to government policies. For example, decreasing income tax rates increases individual's hourly wages. If the labor supply elasticity is negative, hours worked will decrease. If the labor supply elasticity is positive, hours worked will increase. When people work more, more goods and services are produced. Our results suggest that, in 2010, a decrease in income tax rates would cause married males, married females, and unmarried females to work less, while unmarried males would not much change the hours they work.

7. Discussion

In this paper, we have proposed a two-stage least squares estimator for the coefficient of an endogenous functional predictor in a functional concurrent regression model. We showed through simulations that the proposed estimator is far superior to a naive estimation strategy that ignores the endogeneity issue. In addition, we outlined a computational strategy which is easily implemented using existing software. The estimation method is specifically designed to handle cases in which the functional data are observed sparsely, unlike existing methods.

Finally, we applied the FC2SLS estimation strategy to estimate the LSE of both married and unmarried females and males from 1988 to 2014. The estimated labor supply elasticities calculated in Section 6 are meant to be illustrative of the potential for applying functional data methods in general, and our proposed FC2SLS estimation method in particular, to questions of great economic interest. A more careful future analysis might consider adding more instruments. Heim [25], for instance, included

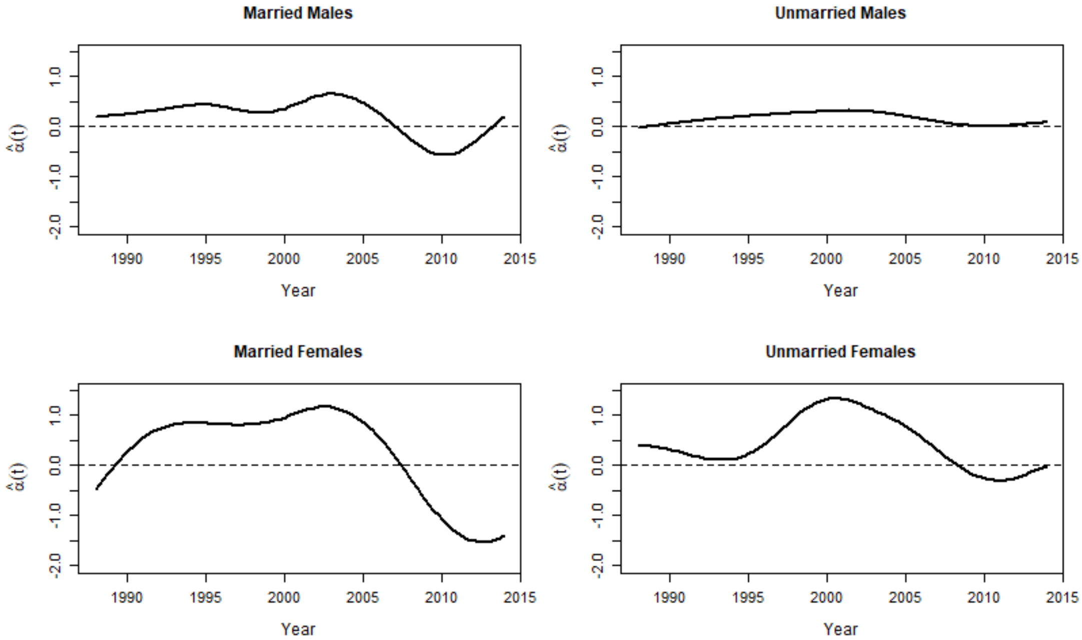

Figure 4. Estimates of labor supply elasticities for married males (upper left), unmarried males (upper right), married females (lower left), and unmarried females (lower right). Dashed horizontal lines at an LSE of 0 are included for reference.

a cubic function of age and of years of education, dummy variables for race, some geographic variables, and the inverse Mills ratio of labor force participation to account for selection bias. The inverse mills ratio measures each individual's propensity to work because many people in the CPS data do not work. Our procedure only applies to those who have chosen to work. Incorporating an additional step in the estimation procedure to calculate the inverse mills ratio would allow us to estimate an elasticity of choosing to participate in the labor market.

Our work also points to many interesting possibilities for future research. For instance, in our simulations, we focus on two different sparse sampling designs. However, it would be useful to know more precisely how the degree of sparsity affects the performance of the FC2SLS estimator, which can be studied both through theory and future simulations. Another issue that is well-studied in the literature on instrumental variable estimation in scalar regression models, but which we did not explore in detail, is the effects of weak instruments. Weak instruments—instruments which are weakly correlated with the endogenous predictors—in scalar regression models are known to exacerbate the bias in small-sample 2SLS estimates and affect statistical inference for the 2SLS estimator [23]. Furthermore, the effects of sparsely-sampled functional data could have interesting interactions with the effects of weak instruments, which should also be investigated in future research. Finally, recent work [see 37, 40] has highlighted the value of treating sparsity in functional data more like a classical missing data problem, opening up a range of imputation methods which better propagate the imputation uncertainty into model estimation. Such methods were not considered here but would enhance future analyses.

As the latest of several papers published in the past 10 years, our work adds to the recently growing literature on instrumental variable estimation in functional regression models. It also showcases one important application of functional data techniques to estimating labor supply elasticities. Our hope is that this inspires the continued adaptation of causal inference methods to functional data settings, which will help to make functional data techniques more useful to economists and many others who commonly grapple with issues of endogeneity.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] D. Aaronson and E. French, *The effects of progressive taxation on labor supply when hours and wages are jointly determined*, The Journal of Human Resources 4 (2009), pp. 386–408.

[2] J.G. Altonji, *Intertemporal substitution in labor supply: Evidence from micro data*, Journal of Political Economy 94 (1986), pp. S176–S215.

[3] J.D. Angrist, *Grouped-data estimation and testing in simple labor-supply models*, Journal of Econometrics 47 (1991), pp. 243–266.

[4] J.D. Angrist, G.W. Imbens, and D.B. Rubin, *Identification of causal effects using instrumental variables*, Journal of the American statistical Association 91 (1996), pp. 444–455.

[5] J.D. Angrist and A.B. Krueger, *Instrumental variables and the search for identification: From supply and demand to natural experiments*, Journal of Economic perspectives 15 (2001), pp. 69–85.

[6] O. Ashenfelter and J. Heckman, *The estimation of income and substitution effects in a model of family labor supply*, Econometrica 42 (1974), pp. 73–85.

[7] A. Babii, *Identification and estimation in the functional linear instrumental regression*, Tech. Rep., UNC Working paper, 2017.

[8] D. Benatia, M. Carrasco, and J.P. Florens, *Functional linear regression with functional response*, Journal of econometrics 201 (2017), pp. 269–291.

[9] F.D. Blau and L.M. Kahn, *Changes in the labor supply behavior of married women: 1980-2000*, Journal of Labor Economics 25 (2007), pp. 393–438.

[10] G.J. Borjas, *The relationship between wages and weekly hours of work: The role of division bias*, Journal of Human Resources 15 (1980), pp. 409–423.

[11] C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler, *Labor supply of new york city cabdrivers: One day at a time*, The Quarterly Journal of Economics 112 (1997), pp. 407–441.

[12] J. Cederbaum, M. Pouplier, P. Hoole, and S. Greven, *Functional linear mixed models for irregularly or sparsely sampled data*, Statistical Modelling 16 (2016), pp. 67–88.

[13] C. Chen, S. Guo, and X. Qiao, *Functional linear regression: dependence and error contamination*, Journal of Business & Economic Statistics 40 (2022), pp. 444–457.

[14] D. Dorn, *Essays on inequality, spatial interaction, and the demand for skills*, Dissertation University of St. Gallen no. 3613 (2009).

[15] Z. Eckstein, M. Keane, and O. Lifshitz, *Career and family decisions: Cohorts born 1935-1975*, Econometrica 87 (2019), pp. 217–253.

[16] P.H. Eilers and B.D. Marx, *Flexible smoothing with b-splines and penalties*, Statistical science 11 (1996), pp. 89–121.
[17] N. Eissa, *Taxation and labor supply of married women: The tax reform act of 1986 as a natural experiment*, NBER Working Paper 5023 (1995).
[18] N. Eissa and H.W. Hoynes, *Taxes and the labor market participation of married couples: the earned income tax credit*, Journal of Public Economics 88 (2004), pp. 1931–1958.
[19] J.P. Florens and S. Van Bellegem, *Instrumental variable estimation in functional linear models*, Journal of Econometrics 186 (2015), pp. 465–476.
[20] A. Gajardo, S. Bhattacharjee, C. Carroll, Y. Chen, X. Dai, J. Fan, P.Z. Hadjipantelis, K. Han, H. Ji, C. Zhu, H.G. Müller, and J.L. Wang, *fdapace: Functional Data Analysis and Empirical Dynamics* (2021). Available at https://CRAN.R-project.org/package=fdapace, R package version 0.5.8.
[21] R. Ghosal and A. Maity, *A score based test for functional linear concurrent regression*, Econometrics and Statistics 21 (2022), pp. 114–130.
[22] J. Goldsmith and J.E. Schwartz, *Variable selection in the functional linear concurrent model*, Statistics in medicine 36 (2017), pp. 2237–2250.
[23] W.H. Greene, *Econometric Analysis*, 8th ed., Pearson, 2018.
[24] J. Heckman, *Shadow prices, market wages, and labor supply*, Econometrica 42 (1974), pp. 679–694.
[25] B.T. Heim, *The incredible shrinking elasticities married female labor supply, 1978–2002*, Journal of Human resources 42 (2007), pp. 881–918.
[26] G. Imbens, *Instrumental variables: an econometrician's perspective*, Statistical Science 29 (2014), pp. 323–358.
[27] A.E. Ivanescu, A.M. Staicu, F. Scheipl, and S. Greven, *Penalized function-on-function regression*, Computational Statistics 30 (2015), pp. 539–568.
[28] G.M. James, T.J. Hastie, and C.A. Sugar, *Principal component models for sparse functional data*, Biometrika 87 (2000), pp. 587–602.
[29] P. Kokoszka and M. Reimherr, *Introduction to functional data analysis*, CRC Press, 2017.
[30] D.R. Kowal, D.S. Matteson, and D. Ruppert, *Functional autoregression for sparsely sampled data*, Journal of Business & Economic Statistics 37 (2019), pp. 97–109.
[31] A. Leroux, L. Xiao, C. Crainiceanu, and W. Checkley, *Dynamic prediction in functional concurrent regression with an application to child growth*, Statistics in medicine 37 (2018), pp. 1376–1388.
[32] A. Leroux, L. Xiao, C. Crainiceanu, and W. Checkly, *fcr: Functional Concurrent Regression for Sparse Data* (2018). Available at https://CRAN.R-project.org/package=fcr, R package version 1.0.
[33] A. Lewbel, *Using instrumental variables to estimate models with mismeasured regressors*, in *Handbook of Measurement Error Models*, G.Y. Yi, A. Delaigle, and P. Gustafson, eds., chap. 5, Chapman and Hall/CRC, 2021, pp. 85–96.
[34] T.E. MaCurdy, *An empirical model of labor supply in a life-cycle setting*, Journal of Political Economy 89 (1981), pp. 1059–1085.
[35] N. Malfait and J.O. Ramsay, *The historical functional linear model*, Canadian Journal of Statistics 31 (2003), pp. 115–128.
[36] J.S. Morris, *Functional regression*, Annual Review of Statistics and Its Application 2 (2015), pp. 321–359.
[37] J. Petrovich, M. Reimherr, and C. Daymont, *Highly irregular functional generalized linear regression with electronic health records*, Journal of the Royal Statistical Society: Series C (Applied Statistics) 71 (2022), pp. 806–833.
[38] J.O. Ramsay and B.W. Silverman, *Applied functional data analysis: methods and case studies*, Springer, 2007.
[39] J.O. Ramsay and B.W. Silverman, *Functional data analysis*, Springer-Verlag, 2005.
[40] A.R. Rao and M. Reimherr, *Modern multiple imputation with functional data*, Stat 10 (2021), p. e331.
[41] D. Şentürk and H.G. Müller, *Functional varying coefficient models for longitudinal data*,

Journal of the American Statistical Association 105 (2010), pp. 1256–1264.
[42] D. Şentürk and D.V. Nguyen, *Varying coefficient models for sparse noise-contaminated longitudinal data*, Statistica Sinica 21 (2011), pp. 1831–1856.
[43] D. Seong and W.K. Seo, *Functional instrumental variable regression with an application to estimating the impact of immigration on native wages*, arXiv preprint arXiv:2110.12722v2 (2022).
[44] C.D. Tekwe, R.S. Zoh, M. Yang, R.J. Carroll, G. Honvoh, D.B. Allison, M. Benden, and L. Xue, *Instrumental variable approach to estimating the scalar-on-function regression model with measurement error with application to energy expenditure assessment in childhood obesity*, Statistics in medicine 38 (2019), pp. 3764–3781.
[45] L. Xiao, C. Li, W. Checkley, and C. Crainiceanu, *Fast covariance estimation for sparse functional data*, Statistics and computing 28 (2018), pp. 511–522.
[46] F. Yao, H.G. Müller, and J.L. Wang, *Functional data analysis for sparse longitudinal data*, Journal of the American statistical association 100 (2005), pp. 577–590.
[47] X. Zhang, Q. Zhong, and J.L. Wang, *A new approach to varying-coefficient additive models with longitudinal covariates*, Computational Statistics & Data Analysis 145 (2020), p. 106912.

Appendix A. Penalization Details for FC2SLS

In this section, we outline in greater detail the regularization described in Section 4 of the main paper. The penalties for each of the two modeling stages are provided below.

(1) *Stage 1*: We first apply a basis function decomposition to the functional intercept as $\gamma_0(t) \approx \sum_{k=1}^{M_0} A_{0,k}(t)\gamma_{0,k}$, where the $A_{0,k}(.)$ are known univariate basis functions and $\gamma_{0,k}$ are the corresponding coefficients. Similarly, we represent the functional coefficients for the functional predictors, $\tilde{X}_p(t)$, using a basis expansion as follows: $\gamma_p(t) \approx \sum_{k=1}^{M_p} A_{p,k}(t)\gamma_{p,k}$, where $A_{p,k}(.)$ is a known univariate basis and $\gamma_{p,k}$ is the corresponding coefficient for $p = 1, ...K_1$. Finally we apply again a basis function decomposition for the functional instrumental variables, $\tilde{Z}_r(t)$, as $\phi_r(t) \approx \sum_{l=1}^{Q_r} B_{r,l}(t)\phi_{r,l}$, where $B_{r,l}(t)$ are known univariate basis functions and $\phi_{r,l}$ are the corresponding coefficients for $r = 1, ...K_3$.

The penalty term for stage 1 can be written as

$$\sum_{i=0}^{K_1} \lambda_i P_i^{(1)}(\boldsymbol{\gamma}_i) + \sum_{r=1}^{K_3} \omega_r P_r^{(2)}(\boldsymbol{\phi}_r),$$

where $P_i^{(1)}(\boldsymbol{\gamma}_i) = \boldsymbol{\gamma}_i^t D_i^{(1)} \boldsymbol{\gamma}_i$ and $P_r^{(2)}(\boldsymbol{\phi}_r) = \boldsymbol{\phi}_r^t D_r^{(2)} \boldsymbol{\phi}_r$, with $\boldsymbol{\gamma}_i$ representing the vector of all $\gamma_{i,k}$, for $i = 0, 1..., K_1$ and $\boldsymbol{\phi}_r$ representing the vector of all $\phi_{r,l}$, for $r = 0, 1..., K_3$. The matrices $D_i^{(1)}$ and $D_r^{(2)}$ are penalty matrices applied to the associated basis and λ_i and ω_r are the scalar smoothing parameters which control the smoothness of the resultant (estimated) coefficient functions.

(2) *Stage 2*: Similarly to stage 1, we use basis function decompositions for the functional intercept and the functional coefficients of the functional predictors, $\tilde{X}_p(t)$ and $\hat{W}(t)$. Applying the basis functions decomposition, these can be expressed as $\beta_0(t) \approx \sum_{k=1}^{L_0} A_{0,k}^*(t)\beta_{0,k}$, $\beta_p(t) \approx \sum_{k=1}^{L_p} A_{p,k}^*(t)\beta_{p,k}$ and $\alpha(t) \approx \sum_{l=1}^{Q} B_l(t)\alpha_l$,

where $A^*_{0,k}(t)$, $A^*_{p,k}(t)$ and $B_l(t)$ are known univariate bases with the corresponding coefficients $\beta_{0,k}$, $\beta_{p,k}$, and α_l.

The penalty term for stage 2 can be written as

$$\sum_{i=0}^{K_1} \lambda_i^* P_i^{(3)}\left(\boldsymbol{\beta}_i\right) + \omega^* P^{(4)}\left(\boldsymbol{\alpha}\right),$$

where $P_i^{(3)}\left(\boldsymbol{\beta}_i\right) = \boldsymbol{\beta}_i^t D_i^{(3)} \boldsymbol{\beta}_i$ and $P^{(4)}\left(\boldsymbol{\alpha}\right) = \boldsymbol{\alpha}^t D^{(4)} \boldsymbol{\alpha}$, with $\boldsymbol{\beta}_i$ representing the vector of all $\beta_{i,k}$, for $i = 0, 1..., K_1$, and $\boldsymbol{\alpha}$ the vector of all α_l. The $D_i^{(3)}$ and $D^{(4)}$ are penalty matrices and λ_i^* and ω^* are the scalar smoothing parameters.

Appendix B. Additional Simulation Results

Table B1 shows the mean integrated square error for all simulations described in Section 5. While the overall findings remain largely the same as those using the median integrated square error, it is evident that for smaller sample sizes, the *FC2SLS-f* approach can suffer from occasional wild estimates. For instance, the mean ISE for $N = 100$ in Scenario 1 is quite large, but drops to a reasonably small level by $N = 200$, and eventually surpasses even the mean ISE of *FC2SLS-p* as the smallest for $N = 400$ and 800. A similar trend is seen for Scenarios 2 and 3 as well. Practically, one may be best advised to use the *FC2SLS-f* approach with caution when using small sample sizes, but with confidence otherwise.

Table B1. Mean integrated squared error of $\hat{\alpha}(t)$ from each method across 100 simulations.

		CSD				CPS			
Scenario	N	Naive-f	Naive-p	FC2SLS-f	FC2SLS-p	Naive-f	Naive-p	FC2SLS-f	FC2SLS-p
1	100	1.03	4.08	16.71	0.19	1.13	4.32	0.40	0.27
	200	1.02	3.93	1.75	0.10	1.06	4.01	0.17	0.19
	400	1.01	3.95	0.06	0.06	1.03	4.05	0.10	0.11
	800	1.00	3.96	0.03	0.03	1.01	4.04	0.04	0.09
2	100	1.72	6.73	33.94	0.51	1.86	7.23	1.21	0.55
	200	1.69	6.47	16.59	0.38	1.73	6.62	0.89	0.42
	400	1.66	6.57	0.90	0.32	1.58	6.56	0.71	0.35
	800	1.65	6.56	0.67	0.30	1.57	6.60	0.65	0.32
3	100	1.44	1.54	5.54	0.28	2.11	1.53	0.90	0.37
	200	1.29	1.38	2.20	0.12	1.93	1.35	0.65	0.23
	400	1.48	1.18	1.14	0.09	1.85	1.14	0.33	0.16
	800	1.39	1.13	0.27	0.06	1.70	0.97	0.27	0.14