Local Cohorts Estimator for Synthetic Panels with Fixed Effects

Lucas Núñez *

Schar School of Policy and Government

George Mason University

May 23, 2023

Abstract

A common method to deal with unobserved heterogeneity in linear models is to rely on fixed effects when panel data is available. However, this is not always the case. In some instances, researchers may have access to Repeated Cross Sectional (RCS) data where a random samples composed of different individuals are taken at different points in time. In this paper I develop a method, called Local Cohorts Estimator, to obtain consistent estimates in linear models with unobserved heterogeneity using RCS data. I show that the method performs well is simulations. Additionally, I include an application using real panel data on media consumption and political views, where I show that LCE can reasonable approximate the results obtained from real panel data.

*Lucas Núñez is an Assistant Professor at the Schar School of Policy and Government, George Mason University. Email: `lnunez6@gmu.edu`. This paper has benefited from feedback received during seminars at Caltech, as well as during the Polmeth conference. I want to thank R. Michael Alvarez, Jonathan Katz, and Robert Sherman for especially helpful advise and recommendations.

1 Introduction

In many circumstances, there is a lack of panel data where specific individuals are followed over time. In many of those situations, however, Repeated Cross-Sectional data (RCS) may be available. In RCS data, a random sample is taken from the same population at several points in time, using the same (or very similar) sampling techniques and questionnaires. A substantial amount of data available for social research, in fact, comes in the form of RCS data; for example, the Current Population Survey (CPS), the Cooperative Congressional Election Study (CCES), the Latin America Public Opinion Project (LAPOP), data from the Pew Research Center, the General Social Survey (GSS), the American National Election Studies (ANES), among others.

The advantage of panel data over RCS data comes precisely from the fact that a single individual is observed at different points in time. This allows researchers to control for unobserved individual-level heterogeneity (by using fixed-effects) that aids in causal identification; or to estimate dynamic models where individuals' past behavior influences their current behavior. In contrast, RCS data has the drawback that it cannot leverage this information, as each individual is observed only once, meaning that standard estimation techniques cannot be applied.

In light of the relative abundance of RCS data compared to panel data, and the need for reliable estimates under the presence of unobserved individual-level heterogeneity, I develop a semi-parametric two-step estimator, the Local Cohorts estimator (LC), that allows for the estimation of linear fixed-effects models using RCS data. I show that, under some assumptions, this estimator produces root-n consistent and asymptotically normal estimates of the parameters of interest. Furthermore, in Monte Carlo simulations, I show that the LC estimator outperforms other available estimators in small samples.

The main assumption on which the LC estimator relies on is that similar individuals have, on average, similar unobserved individual fixed-effects, where similarly is measured with respect to a set of observable time-invariant individual-level characteristics. Under this assumption, the first step of the LC estimator generates a synthetic panel, where each observation corresponds to the average individual with some time-invariant characteristics observed in the RCS data. In the second step, the LC estimator applies the within transformation, common in fixed-effects models, to the synthetic panel generated in the first step to obtain estimates of the parameters of interest.

The LC estimator has advantages over other estimators that have been proposed for fixed-effects models using RCS data. These estimators fall in two categories. The first, and closest to the LC estimator, is the cohort averaging approach developed by Deaton (1985). In this approach, individuals sharing common characteristics (most typically gender and year of birth) are grouped into disjoint cohorts, after which averages within these cohorts are treated as observations in a synthetic panel.[1] This stands in contrast with the LC estimator, which allows observations in the RCS data to belong to multiple groups, being weighted within each group depending on how closely they share observable time-invariant characteristics with other members of the group. This allows the LC estimator to make better use of the information available in the RCS data, compared to the arbitrary splitting of the data in Deaton's approach. The second approach, first developed by Moffitt (1993), relies on observable time-invariant characteristics and functions of time to instrument for variables in the model that are correlated (or suspected to be) with the unobserved fixed effects.[2]

Both these approaches suffer from some problems. The cohort averaging approach can produce very biased estimates in small samples, even after using bias-correction techniques. This is, in large part, because it cannot resolve the tension that exists between the cohorts sizes and the number of cohorts: for this estimator's assumptions to be tenable, it necessitates a large number of observations per cohort, but this necessarily implies that the number of cohorts be small, which translates into a synthetic panel with very few observations.[3] Another shortcoming is that the design of the cohorts is entirely up to the researcher, which can give rise to different estimates from the same data by two researchers seemingly using the same estimation method.

[1] See also, Verbeek and Nijman (1992); Collado (1997); Devereux (2007); McKenzie (2004); Ridder and Moffitt (2007), among others.

[2] See also, Ridder and Moffitt (2007); Pelser et al. (2002, 2004), among others. Verbeek (2008) provides a short review of the literature on the cohort averaging and instrument variable approaches.

[3] Of course, if the data set contains hundreds of thousands of observations, then this tension is not a serious constraint.

The instrumental variables (IV) approach can suffer from the weak instruments problem, with a first step estimation that has too little variance. This then translates into a large degree of uncertainty around the estimates of the parameters of interest. This approach also requires a correct specification of the instruments' equation which, in general, is unknown. Therefore, a workable version of the IV approach requires the use of a flexible first step specification that allows for an unknown functional form. While this approach can generate consistent estimates of the parameters of interest, using a flexible specification in the first step estimation necessarily leads to larger uncertainty around the estimates of the parameters of interest.

The Local Cohorts estimator is better able to resolve the tension between cohort size and the number of cohorts, compared to Deaton's approach, as it allows observations from the RCS data to belong to multiple cohorts. This translates into smaller bias in small samples but, because of the greater complexity of the first step estimation, it can lead to a larger variance. My simulations results suggest that the reduction in bias is greater than the increase in the variance, leading to a smaller Root Mean Squared Error (RMSE) for the LC estimator. Another advantage of the LC approach is that the non-parametric nature of the first step better captures non-linearities in the relation between the unobserved fixed-effects and the observed time-invariant characteristics that are used to construct the cohorts. This also contributes to reducing the bias of the estimator in small samples. Despite these benefits, the Local Cohorts approach, as well as Deaton's approach, suffer from the curse of dimensionality, as sample size can restrict the number of time-invariant characteristics that can be used to construct the cohorts.[4] This issue is reduced in the case of Moffitt's approach. To the extent that the fixed-effect is reflecting highly multidimensional individual-level heterogeneity, Moffitt's approach could be superior.

Additionally, I study the performance of the LC estimator relative to a fixed-effects estimator from actual panel data. The effect of interest is the impact of partisan news exposure on the favorability ratings of Barack Obama. In this study, I find that the LC estimator outperforms

[4] This problem could be partially avoided by resorting to dimension-reducing techniques, like principal components.

alternative RCS data estimators. Moreover, while there is some amount of bias relative to the actual panel estimates, the results show that the identification assumptions of the LC estimator plausibly hold in this case.

The rest of this paper is organized as follows. In Section 2 I present the Local Cohorts estimator and I show that it is root-n consistent and asymptotically normally distributed. In Section 3 I describe Deaton's and Moffitt's approaches to the estimation of fixed-effects models with RCS data. In Section 4 I compare the performance of the LC estimator with alternative estimators under different specifications, using Monte Carlo simulations. In Section 5 I estimate the effects of partisan news exposure on voters' perceptions of Barack Obama using the LC estimator, finding them comparable to those derived from actual panel data. Finally, in Section 6 I conclude and suggest avenues for further research.

2 Unobserved Heterogeneity and the Local Cohorts Estimator

Before presenting the estimation of models with unobserved individual-level heterogeneity using RCS data, it is useful to review the estimation of these models when panel data is available, as this serves as the basis for the Local Cohorts estimator developed in this paper. Consider the following model with unobserved individual-level heterogeneity (fixed-effects):

$$y_{it} = x_{it}\beta_0 + f_i + \varepsilon_{it}, \quad t = 1, ..., T, \quad i = 1, ..., n. \tag{1}$$

where x_{it} denotes a k-dimensional vector of real-values explanatory variables, $\beta_0 \in \mathbb{R}^k$ are the parameters of interest, $f_i \in \mathbb{R}$ are individual-level unobserved effects that may be corelated with x_{it}, and ε_{it} is an error term with mean 0 and variance σ^2 such that $E(\varepsilon|x, f) = 0$. Because x_{it} and f_i are (potentially) correlated, estimating β_0 via an OLS regression of y_{it} on x_{it} ignoring f_i yields biased and inconsistent estimates. When panel data is available, this problem can be solved by applying the within transformation to the model in equation 1, which demeans the variables for each individual. Since the unobserved individual-level heterogeneity is assumed to be constant in time, demeaning eliminates the unobserved term f_i from equation 1, and β_0 can be then consistently estimated via OLS on the transformed equation.

When only RCS data is available, however, each individual in the sample is only observed at a single time period (or cross-section). That is, letting i_t denote that individual i is observed at time t only, the data generating process for the model with unobserved individual-level heterogeneity can be re-written as:

$$y_{i_t t} = x_{i_t t}\beta_0 + f_{i_t} + \varepsilon_{i_t t} \tag{2}$$

Precisely because in RCS data each individual is only observed at only one time period, the within transformation cannot be applied to eliminate the individual-level heterogeneity, f_{i_t}

2.1 Local Cohorts Estimator

Consider the model in equation 2 and a RCS sample of $(y_{i_t t}, x_{i_t t}, z_{i_t})$, where $y_{i_t t} \in \mathbb{R}$ is the outcome of interest, $x_{i_t t} \in \mathbb{R}^k$ are the explanatory variables, and $z_{i_t} \in \mathcal{Z} \subset \mathbb{R}^d$ are additional observable individual-level characteristics. Calculating the expectation of equation 2 conditional on a given value z_c of the observed time-invariant characteristics results in:

$$E(y_{i_t t}|Z = z_c) = E(x_{i_t t}|Z = z_c)\beta_0 + E(f_{i_t}|Z = z_c) + E(\varepsilon_{i_t t}|Z = z_c) \tag{3}$$

which, by denoting $r_{ct} = E(r_{i_t t}|Z = z_c)$ for any variable r, can be written as:

$$y_{ct} = x_{ct}\beta_0 + f_{ct} + \varepsilon_{ct} \tag{4}$$

Under the assumption that conditional on a value of z, the expected fixed-effects from different cross-sections are the same, that is, $f_{ct} = f_c \; \forall t$, equation 4 reduces to:

$$y_{ct} = x_{ct}\beta_0 + f_c + \varepsilon_{ct} \tag{5}$$

By using multiple values of the time-invariant characteristics, it is possible to generate multiple (synthetic) observations as the one in equation 5. That is, computing equation 3 for values $(z_1, ..., z_C)$, with $z_c \in \mathcal{Z}$, results in a synthetic panel of size $C \times T$, where I refer to each $c = 1, ..., C$ as a local cohort. In order to maintain the sampling properties of the original sample, the values of $(z_1, ..., z_C)$ should be selected in a way that maintains the distribution of these observed time-invariant characteristics in the synthetic panel similar to those of the underlying population (or the original sampling design). This can be easily achieved by using as conditioning values of z those actually observed in the RCS data. For example, those from the first cross-section: $z_{i_t} : i_1 = 1, ..., n$.

Based on the synthetic panel generated as in equation 3, it is possible to estimate β_0 by applying OLS on the within transformation of equation 4. Thus, I define the Local Cohorts estimate of β_0 as:

$$\hat{\beta}_{LC} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{Y} \tag{6}$$

where $\widetilde{Y}$ is the matrix form of $\widetilde{y}_{ct} = y_{ct} - \sum_{t=1}^{T} \frac{1}{T} y_{ct}$, and similarly for $\widetilde{X}$.

The challenge of the procedure described in equation 3 is to obtain sample estimates of the conditional expectations that are used to create the synthetic panel. They key assumption required for identification is that $E(f_{i_t}|Z = z_c) \equiv f_{ct} = f_c, \; \forall t$; that is, individuals who are similar on observables, have on average the same unobserved individual-level heterogeneity. For this assumption to hold in practice, it is necessary for each local cohort to have a sufficiently large number of observations. In that case, it is reasonable to say that the average of individuals with the same observed time-invariant characteristics has the same average fixed-effect at different time periods. Given that the idea is to generate a local cohort for each individual observed in the first cross-section of the RCS data, each local cohort at each point in time will typically contain only one individual such that $z_{i_t} = z_c$.[5] Thus, using simple averaging to approximate the conditional expectations will not work, as it is unjustifiable to say that the average fixed-effect for each cohort to be constant across time with only one observation in each local cohort.

To solve this issue, I use non-parametric estimation to generate the synthetic panel. The idea is that individuals in a neighborhood of z_c also are similar in terms of fixed-effects. Hence, in order to capture the average fixed-effect for the local cohort defined by z_c, I average the observations in the neighborhood of z_c. For this approach to be valid, it requires a slight modification of the fixed-effects stability assumption described earlier. In particular, it requires that $E(f_{i_t}|Z = z_c) = \psi(z_c)$, where $\psi(\cdot)$ is a continuous function of z_c. That is, it requires that, on average, the fixed-effect of individuals who are similar in terms of their observable time-invariant characteristics, be also similar, with the relation being stable across time.

More formally, I propose a two-step semi-parametric estimator of the parameter of interest, β_0, which I call the Local Cohorts estimator (LC). In the first (non-parametric) step, the conditional expectations of $y_{i_t t}$ and $x_{i_t t}$ on z_c (denoted y_{ct} and x_{ct}) are obtained using kernel estimators:

$$y_{ct} = \frac{\sum_{i_t=1}^{n_t} \mathcal{K}\big(S^{-1/2}(z_{i_t} - z_c)/h\big) y_{i_t t}}{\sum_{i_t=1}^{n_t} \mathcal{K}\big(S^{-1/2}(z_{i_t} - z_c)/h\big)}$$

and

$$x_{ct} = \frac{\sum_{i_t=1}^{n_t} \mathcal{K}\big(S^{-1/2}(z_{i_t} - z_c)/h\big) x_{i_t t}}{\sum_{i_t=1}^{n_t} \mathcal{K}\big(S^{-1/2}(z_{i_t} - z_c)/h\big)}$$

where $\mathcal{K}(u)$ denotes a kernel function, h is a bandwidth, and S is the variance-covariance matrix of the time-invariant covariates z.[6]

The second (parametric) step uses the estimated values y_{ct} and x_{ct} for multiple values of z_c as the input in a within estimator for fixed-effects. That is, I estimate the finite dimensional

[5] Except when z takes on discrete values.

[6] This is exactly the same specification as the multidimensional Naradaya-Watson non-parametric regression estimator.

parameter vector β_0 by:

$$\widehat{\beta}_{LC} = \arg\min_{\beta \in \mathcal{B}} \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{1}{2} (\widetilde{y}_{ct} - \beta \widetilde{x}_{ct})^2 \tag{7}$$

where $\widetilde{y}_{ct} = y_{ct} - \frac{1}{T} \sum_{t=1}^{T} y_{ct}$ is the within transformation of y_{ct}, and similarly for $\widetilde{x}_{ct}$.

While any choice of z_c values is valid, I advocate the use of those z actually observed in the first-cross section of the RCS data to generate the synthetic panel. This has two practical advantages relative to arbitrary user-specified values. First, using values of z actually observed in the sample ensures that the estimates incorporate the sample distribution of these covariates, ensuring a representative sample.[7] Second, it reduces the amount of extrapolation necessary in the first step estimates, as it guarantees that there exists at least one observation with said value.

2.2 Asymptotic Properties of the Local Cohorts Estimator

To establish the asymptotic properties of the Local Cohorts estimator, it is useful to define some notation first. Let $(\widetilde{Y}, \widetilde{X})$ denote the matrix form of the within transformation of the estimates from the first step, $(\widetilde{Y}_0, \widetilde{X}_0)$ their population values; finally, $G_n(\beta, \widetilde{Y}, \widetilde{X}) = \sum_{c=1}^{C} \sum_{t=1}^{T} g(\beta, \widetilde{y}_{ct}, \widetilde{x}_{ct})$, where $g(\beta, \widetilde{y}_{ct}, \widetilde{x}_{ct}) = \frac{1}{2}(\widetilde{y}_{ct} - \beta \widetilde{x}_{ct})$, the squared loss function. Notice that both $(\widetilde{Y}, \widetilde{X})$ and $(\widetilde{Y}_0, \widetilde{X}_0)$ are functions of z.

In what follows, I first formally define the assumptions I use to establish the asymptotic properties of the estimator and then discuss their meaning and the role they play in the proofs.

A1. $E(\varepsilon | z, x) = 0$, with ε i.i.d. with mean 0 and variance 1.

A2. For all t, $E(f_{i_t} | Z = z_c) = \psi(z_c)$, where $\psi(\cdot)$ is a continuous function of z_c.

A3. $|\mathcal{K}(u)| < \infty$, and $\int_{\mathbb{R}^d} |\mathcal{K}(u)| du < \infty$, and $\mathcal{K}(u)$ is symmetric, and $\int |u|^2 |\mathcal{K}(u)| du < \infty$.

[7] This relies on the implicit assumption, throughout the paper, that the RCS data are representative samples from the same underlying population.

A4. For some $\Lambda_1 < \infty$ and $L < \infty$, either $\mathcal{K}(u) = 0$ for $||u|| > L$ and for all $u, u' \in \mathbb{R}^d$, $|\mathcal{K}(u) - \mathcal{K}(u')| \leq \Lambda_1 ||u - u'||$, or $\mathcal{K}(u)$ is differentiable, $|\frac{\partial}{\partial u}\mathcal{K}(u)| \leq \Lambda_1$, and for some $\nu > 1$, $|\frac{\partial}{\partial u}\mathcal{K}(u)| \leq \Lambda_1 ||u||^{-\nu}$, for $||u|| > L$.

A5. $f(z)$, $f(z) \times E(y|Z = z)$, and $f(z) \times E(x|Z = z)$ are uniformly continuous and bounded.

Assumption A1 is simply a strict exogeneity assumption, common in regression analysis, and I assume the variance of ε to be one to simplify notation.[8] Assumption A2 is the key identifying assumption. It states that, on average, the fixed-effects of individuals who are similar to each other in terms of their observable time-invariant characteristics, is similar. Assumption A3 imposes some mild restrictions on the kernel $\mathcal{K}(\cdot)$. Assumption A4 requires that the kernel $\mathcal{K}(\cdot)$ be sufficiently smooth, in this case that it either has truncated support and is Lipschitz continuous, or that it has a bounded derivative with an integrable tail.[9] Assumptions like A4 are common in the literature of semi-parametric two-step M-estimators.[10] Finally, assumption A5 requires that the expectations of y and x conditional on z be continuous.

To show the consistency of the Local Cohorts estimator, it is first useful to establish the following uniform convergence in probability result:

Proposition 1 (Uniform Convergence)**.** *Under assumptions A1-A5, and $h \to 0$ "fast enough,"*

$$\sup_{\beta \in \mathcal{B}} |G_n(\beta, \widetilde{Y}, \widetilde{X}) - E(g(\beta, \widetilde{Y}_0), \widetilde{X}_0)| \xrightarrow{p} 0$$

Proof. See Appendix.

The proof of this proposition relies on using the triangle inequality to separate the previous equation into a term that depends on the estimated functions of the conditional expectation of y and x with respect to z, $(\tilde{Y}, \tilde{X})$, and a term that depends on the population functions of the

[8]This could be weakened to $E(\varepsilon[x\ z]) = 0$.

[9]Hansen (2008)) notes that most commonly used kernels satisfy this assumption, including the polynomial kernel class, the higher order polynomial kernels of Muller (1984) and Granovsky and Muller (1991), the normal kernel, and the higher order Gaussian kernels of Wand and Schucany (1990) and Marron and Wand (1992).

[10]See, for example, Newey (1994); Escanciano et al. (2012, 2014); Mammen et al. (2012); Hahn and Ridder (2013).

conditional expectations, $(\tilde{Y}_0, \tilde{X}_0)$. The latter term can be shown to converge in probability to zero by standard results for M-estimators (see, for example, Newey and McFadden, 1994) as it does not depend on the kernel estimates. The former term relies on uniform convergence of kernel estimators coupled with the smoothness of $g(\cdot)$ to prove that it also converges in probability to zero.[11] The requirement that $h \to 0$ "fast enough" means that $O_p((\frac{\ln(n)}{nh^d})^{1/2} + h^2) = o_p(1)$; this requirement is not particularly stringent, and it is satisfied by the optimal bandwidth for the Naradaya-Watson non-parametric regression estimator, $h \propto n^{-1/(d+4)}$.[12]

Proposition 2 (Consistency)**.** *Under assumptions A1-A5, and $h \to 0$ "fast enough," the Local Cohorts estimator $\widehat{\beta}_{LC}$ is consistent:*

$$\widehat{\beta}_{LC} \xrightarrow{p} \beta_0$$

Proof. See Appendix.

The proof of this proposition relies on the consistency of M-estimators (see, for example, Theorem 2.1 in Newey and McFadden (1994). Assumptions A1 and A2 provide the identification conditions, so that the population parameter vector β_0 maximizes the population objective function $E(g(\beta, \widetilde{Y}_0, \widetilde{X}_0))$. Proposition 1 establishes that the sample analog of the objective function is asymptotically arbitrarily close to its population version, so that the maximum of $G_n(\beta, \widetilde{Y}, \widetilde{X})$ converges to the maximum of $E(g(\beta, \widetilde{Y}_0, \widetilde{X}_0))$, which is β_0.

Proposition 3 (Asymptotic Normality)**.** *Under assumptions A1-A5, and $h \to 0$ "fast enough,"*

$$\sqrt{nT}\left(\hat{\beta}_{LC} - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, V_0^{-1}\Sigma_0 V_0^{-1})$$

where

$$V_0 = \frac{\partial^2}{\partial \beta^2} g(\beta_0, \widetilde{Y}_0, \widetilde{X}_0) = \widetilde{X}_0' \widetilde{X}_0$$

and

$$\Sigma_0 = E(\Gamma_0' \Gamma_0)$$

[11] See, for example, Andrews (1995); Fan and Yao (2003); Hansen (2008).

[12] See, for example, Pagan and Ullah (1999), Chapter 3.

where

$$\Gamma_0(z) = \left[\frac{\partial}{\partial\beta} g(\beta_0, \widetilde{Y}_0, \widetilde{X}_0)\right] + \sum_{v=\widetilde{y},\widetilde{x}_1,..,\widetilde{x}_d} D_v \left[\frac{\partial}{\partial\beta} g(\beta_0, \widetilde{Y}_0, \widetilde{X}_0)\right]\left[v - v_0\right]$$

where D_v denotes the first-order derivative with respect to $v = \widetilde{y}, \widetilde{x}_1, ..., \widetilde{x}_k$.

Proof. See Appendix.

The proof of this proposition comes from verifying the conditions for Theorem 3.2 in Ichimura and Lee (2010). V_0^{-1} is the standard variance estimator for the case in which the regressors used in the second step are known. Σ_0 incorporates the effects of the first (non-parametric) step estimation in the final estimate for β_0. Notice that if there were no first step estimation, the second term in $\Gamma_0(z)$ would be zero, and then Σ_0 would simplify to $\widetilde{X}_0'\widetilde{X}_0$. In that case, the overall variance of the estimator simplifies to $(\widetilde{X}_0'\widetilde{X}_0)^{-1}$, which is the standard variance for the fixed-effects estimator when the regressors are known (instead of being generated by the first step of the estimation).

3 Alternative Estimators

3.1 Cohort Averaging Approach

The first approach to estimating fixed-effects models with RCS data is the cohort averaging approach of Deaton (1985). This approach uses cohorts to generate a synthetic panel that is then used to estimate β_0, similarly to the Local Cohorts estimator. In Deaton's approach cohorts are defined as a partition of space of observed time-invariant characteristics $\mathcal{Z} = \bigcup_{c=1}^{C} Z_c$, where $Z_c \cap Z_{c'} = \emptyset$ for all c and c'. For example, if $\mathcal{Z}$ includes gender and year of birth, cohorts in this approach may be defined as disjoint groups by gender and decade of birth.[13] Deaton then averages the RCS observations by cohort and time, obtaining the following equation based

[13] Browning et al. (1985) use cohorts of households defined on the basis of five year bands and whether the heads of the household is a manual or non-manual worker; Blundell et al. (1994b) use year of birth intervals of 10 years, interacted with two education groups; Banks et al. (1994) use five year age groups; Propper et al. (2001) use seven birth groups and ten regions to build their cohorts.

on equation 2:

$$\bar{y}_{ct} = \bar{x}_{ct}\beta_0 + \bar{f}_{ct} + \bar{\varepsilon}_{ct} \tag{8}$$

where $\bar{y}_{ct} = \sum_{i_t=1}^{n} y_{i_t t}\mathbb{I}(z_{i_t} \in Z_c)/\sum_{i_t=1}^{n} \mathbb{I}(z_{i_t} \in Z_c)$, the average of the $y_{i_t t}$s that belong to cohort c observed in cross-section t, and similarly for $x_{i_t t}$, f_{i_r} and $\varepsilon_{i_t t}$. This results in a synthetic panel with observations on C cohorts across T periods. To estimate β_0 from equation 8, Deaton assumes that $\bar{f}_{ct} = f_c$ for all c, which implicitly states that the unobserved individual-level heterogeneity actually has the form of group effects (with some noise around each group). This assumption is related to assumption A2 in the Local Cohorts estimator, but it uses a coarser aggregation. While the continuity assumption in the Local Cohorts estimator establishes a local restriction, Deaton's assumption is about larger sets. Under that assumption, one can apply the within transformation on equation 8 and obtain an estimate of β_0 by OLS:

$$\widehat{\beta}_D = \left(\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct} - \bar{x}_c)(\bar{x}_{ct} - \bar{x}_c)'\right)^{-1}\left(\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct} - \bar{x}_c)(\bar{y}_{ct} - \bar{y}_c)\right) \tag{9}$$

where $\bar{x}_c = \sum_{t=1}^{T} \frac{1}{T}\bar{x}_{ct}$, and similarly for $\bar{y}_c$.

The asymptotic behavior of this estimator can be obtained using several alternative asymptotic sequences; this is because in addition to the two dimensions in panel data (n and T), there are two other dimensions in cohort models: the number of cohorts C and the size of the cohorts n_c.[14]

Deaton (1985) also proposes a related estimator that has a lower dependence on the number of observations per cohort tending to infinity (which is needed to ensure that $\bar{f}_{ct} = f_c$). Considering the cohort averages $\bar{y}_{ct}$ and $\bar{x}_{ct}$ to be measurements of population values $y^{\star}_{ct}$ and $x^{\star}_{ct}$ with errors, he proposes to use an errors-in-variables model, in which the measurement errors

[14]Deaton (1985); Verbeek and Nijman (1993); Collado (1997) assume that the number of cohorts tends to infinity, with cohort sizes held (roughly) constant (which implies that the number of individuals tends to infinity). Moffitt (1993); Verbeek and Vella (2005) consider the case in which the number of individuals tends to infinity while the number of cohorts is held fixed (thus cohort sizes tend to infinity). McKenzie (2004) considers the case in which $T \to \infty$ and the cohort sizes tend to infinity.

are distributed with mean zero and independently of the true values; that is:

$$\begin{pmatrix} \bar{y}_{ct} \\ \bar{x}_{ct} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{pmatrix} y^{\star}_{ct} \\ x^{\star}_{ct} \end{pmatrix} ; \begin{pmatrix} \sigma_{00} & \sigma' \\ \sigma & \Sigma \end{pmatrix} \right) \tag{10}$$

With estimates of Σ and σ in equation 10, the within estimator can be adjusted to eliminate the variation due to measurement error:

$$\hat{\beta}_{DE} = \left(\sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c)(\bar{x}_{ct} - \bar{x}_c)' - \tau \hat{\Sigma} \right)^{-1} \left(\sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c)(\bar{y}_{ct} - \bar{y}_c) - \tau \hat{\sigma} \right) \tag{11}$$

where $\hat{\Sigma}$ and $\hat{\sigma}$ are estimators of Σ and σ that can be obtained from the data, and $\tau = 1$ in Deaton (1985), but Verbeek and Nijman (1993) show that $\tau = (T-1)/T$ has better small sample properties.[15]

The cohort averaging approach to estimating fixed-effects models with RCS data relies on the assumption that, on average, the fixed-effect for a given cohort is the same for all cross-sections under analysis. In practical terms, this requires that in each cohort there is a sufficient number of observations such that the average is somewhat close to the population value, even when using the estimator in equation 11. That is, each cohort needs to average across a reasonably large number of observations from the the RCS data. In empirical applications, Browning et al. (1985) use cohort sizes of about 190 individuals, while Blundell et al. (1994a) use cohort sizes of around 500. More recently, Devereux (2007) argues that cohort sizes should be much larger than that, possibly over 2,000 individuals. The need for large cohorts puts a strain on the data, as larger cohorts necessarily imply fewer cohorts overall, which translates into a synthetic panel with few observations. This may not be a very strong restriction for the CPS data used in many of the empirical applications of this method, as cross-sectional sizes are around 150,000 individuals. But for smaller sample sizes, it creates a tension between the bias generated by small cohort sizes and the uncertainty over the estimates generated by having few observations

[15] For more details on this estimator, please see Deaton (1985); Verbeek and Nijman (1993); Ridder and Moffitt (2007).

in the synthetic panel.

Another significant shortcoming of the cohort averaging approach is that the way cohorts are constructed is important. A priori, there is no guideline as to how to define a cohort, which can lead to different researchers working with the same data to use completely different cohorts and obtain different estimates of the quantities of interest. Besides the arbitrariness issue, Deaton's cohorts can also lead to some groupings that are not entirely reasonable. For example, suppose cohorts are defined by gender and decade of birth. One such cohort might comprise men born between 1960 and 1969, and another men born between 1970 and 1979. This way of constructing cohorts implies that a man born in 1969 has more in common (in terms of his unobserved individual-level characteristics) with a man born in 1960 than with one born just a year after him, in 1970. The use of disjoint cohorts always allows for unreasonable groupings such as this one, regardless of the specifics how cohorts are defined.

The Local Cohorts estimator developed in this paper significantly reduces the shortcomings of the cohort averaging approach. First, by using a local definition of cohort and by allowing observations to belong to multiple cohorts, it avoids the tension between cohort size and the number of cohorts that is inherent to Deaton's approach. This implies that the Local Cohorts estimator is better able to capture the underlying unobserved individual-level heterogeneity, especially in smaller samples where this tension is more patent. Second, by defining cohorts based on neighborhoods whose size depends on the characteristics of the data and can overlap with each other, the Local Cohorts estimator avoids the arbitrariness that cohorts in Deaton's approach have.

3.2 Instrumental Variables Approach

Moffitt (1993) proposes to use instruments to solve the omitted variable bias that comes with the unobserved fixed-effects. As Deaton's approach, it relies on a vector of observed time-invariant individual-level characteristics $z_{i_t} \in \mathcal{Z} \subset \mathbb{R}^d$. Moffitt also considers a vector $w_{i_t t} \in \mathcal{W} \subset \mathbb{R}^m$ of time-varying variables that are uncorrelated with the fixed effects f_{i_t}. These variables $w_{i_t t}$ may simply consist of functions of t. Moffitt's IV approach is based on the following two equations:

$$x_{i_t t} = \delta_1 w_{i_t t} + \delta_2 z_{i_t} + e_{i_t t} \tag{12}$$

and

$$f_{i_t} = z_{it}\gamma + \nu_{i_t} \tag{13}$$

To estimate β_0 with this approach, one first obtains the predicted values $\widehat{x}_{i_t t}$ from equation 12, as in any IV approach, and the estimates the following equation via OLS:

$$y_{i_t t} = \widehat{x}_{i_t t}\beta + z_{i_t}\gamma + \nu_{i_t} + \varepsilon_{i_t t} \tag{14}$$

Letting y, X, Z and ν be the stacked $n \times T$ vectors for all i and t, and defining $U = [X\ Z]$ and $\widehat{U} = [\widehat{X}\ Z]$, the IV estimator of β_0 is consistent if $\text{plim}\frac{1}{nT}\widehat{U}'\nu = 0$ (and $\widehat{U}$ is of full rank, $d+m$), which can be achieved as n goes to infinity, holding T fixed. Notice that the assumption that $\text{plim}\frac{1}{nT}Z'\nu = 0$ is similar to Deaton's assumption that the average of the individual fixed-effects is time invariant. In fact, if z is defined as cohort dummies interacted with time, Moffitt's and Deaton's estimators are identical (see Verbeek, 2008).

One of the advantages of Moffitt's approach is that it permits "grouping" conditional on any set of time-invariant characteristics, whereas cohort approaches are limited by the curse of dimensionality. Another advantage is that it allows for the inclusion of extra time-varying covariates that are uncorrelated with the fixed effects. This allows for greater variation in the outcomes of the first stage IV, that can increase the precision of the estimator.

A clear shortcoming of Moffitt's approach is that it requires a correct specification of the first stage model. For example, if the fixed-effect is related to the observed time-invariant characteristics via a quadratic function, but the model is specified with a linear function, the assumption that $Z'\nu$ approaches zero as n goes to infinity will not hold. This problem can potentially be addressed by using a flexible functional form in the first stage estimation, but this incurs in efficiency costs. Moreover, despite the attractiveness of using time-varying covariates,

$w_{i_t t}$, that are unrelated to the fixed-effects, determining whether a variable is unrelated to the fixed-effect is hard to do.

4 Monte Carlo Simulations

To study the small sample properties of the Local Cohorts estimator, I conduct a series of Monte Carlo simulations and analyze the bias, standard error, and root mean squared error (RMSE) of the Local Cohorts estimator, Moffitt's IV estimator, and two versions of Deaton's cohort averaging approach with different cohort sizes.[16]

For these simulations I consider the following data generating process:

- $y_{it} = 2x_{it} + f_i + \varepsilon_{it},\ \varepsilon_{it} \sim \mathcal{N}(0, 10)$
- $x_{it} = \frac{1}{2}f_i^2 + f_i + \nu_{it},\ \nu_{it} \sim \mathcal{N}(0, 5)$, so that x and f have a non-linear relation.

I consider three cases in terms of the relation between the time-invariant characteristics z and the fixed-effect f:

- **Base Case:** $\mathcal{Z} \subset \mathbb{R}^2,\ z_1 \sim \mathcal{U}(-15, 15),\ z_2 \sim \mathcal{N}(0, 2)$, and $f_i = z_{1i}\sin(z_{1i}/6) + z_{2i} + \eta_i,\ \eta_i \sim \mathcal{N}(0, 2)$.
- **Group Effects Case:** $\mathcal{Z} \subset \mathbb{R}^2,\ z_1 \sim \mathcal{U}(-15, 15),\ z_2 \sim \mathcal{N}(0, 2)$, and $f_i = \frac{1}{2}\left(\sum_{q=1}^{5} \mathbb{I}(z_{1i} \in q(z_1)) + \sum_{q=1}^{5} \mathbb{I}(z_{2i} \in q(z_2))\right)$, where q denotes the quintiles of z and $q(z)$ is the set that indicates that an observation belongs to the qth quintile. Therefore, in this case, f is simply the average of the quintiles to which the observation belongs to in terms of z_1 and z_2, taking values on the set $\{1, \frac{3}{2}.2, \frac{5}{2}, 3, \frac{7}{2}, 4, \frac{9}{2}, 5\}$.
- **Underspecified Case:** $\mathcal{Z} \subset \mathbb{R}^3,\ z_1, z_2, z_3 \sim \mathcal{N}(0, 2)$, with $corr(z_1, z_2) = corr(z_1, z_3) = 0.2$ and $corr(z_2, z_3) = -0.3$, with z_3 unobserved, and $f_i = z_{1i}\sin(z_{1i}/6) + z_{2i} + z_{3i} + \eta_i,\ \eta_i \sim$

[16] For Moffitt's estimator I use a polynomial in z for equations 12 and 13 to capture the non-linearities in the data generating process for both x and f with respect to z. For Deaton's estimator I use versions that have cohorts of approximate size 50 and 200. These cohorts are created by splitting the data along percentiles of the observable time-invariant covariates such that each cohort has (approximately) the same size.

$\mathcal{N}(0, 2)$.[17]

The Base Case satisfies all the assumptions for consistency and asymptotic normality of the Local Cohorts estimator. The Group Effects Case satisfies the assumptions in Deaton's estimator, but represents a mild violation of the continuity assumption (A2) of the Local Cohorts Estimator. The violation is mild because it only occurs at a finite number of points (16). This case is included to show that, even when the continuity assumption does not hold, the Local Cohorts estimator can still perform relatively well. The Underspecified Case incorporates an additional time-invariant characteristic that affects the fixed-effect and is correlated with the other time-invariant characteristics but that is not observed. This case is designed to show that an incomplete specification of the first step model does not affect the Local Cohorts estimator's main properties, but it compromises the validity of Moffitt's approach.[18]

For all cases, I draw samples for two time periods, $T = 2$, with cross-sectional sizes between 250 and 5,000, in increments of 250. Unless otherwise noted, for each sample size and case, I draw a total of 200 Monte Carlo samples.

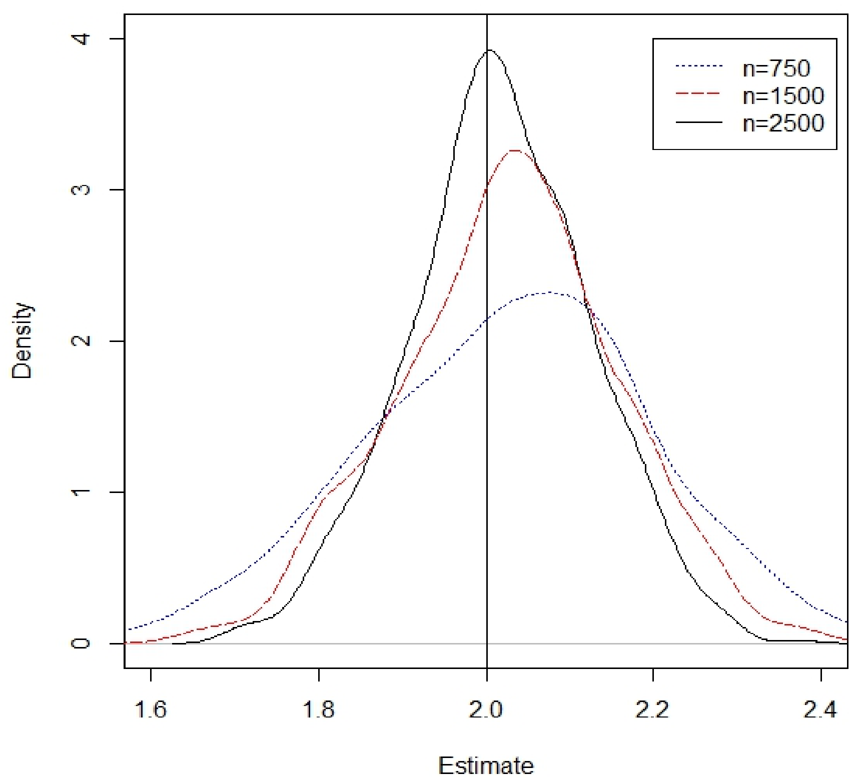
Figure 1 shows 1,000 estimates from the Local Cohorts estimator for the Base Case with three different cross-sectional sample sizes: 750, 1,500, and 2,500. As can be seen from the figure, when the sample size is relatively small the distribution of the estimates is not centered around the true value of the coefficient (in this case, β_0=2), and the distribution is skewed and is slightly platykurtic. When the sample size is larger, like in the case with 2,500 observations per cross-section, the distribution of the estimates looks centered at the true value of the coefficient, and has almost no skewness and is mesokurtic, meaning that for moderate samples sizes a normal distribution describes the distribution of the estimates sufficiently well.

Figure 2 (and Table B1 in the Appendix) shows the Monte Carlo simulation results for the

[17] The distribution of z_1 in the Underspecified case is different from the distribution used for this variable in the other cases. This was done to simplify the creation of non-independent variables. If the simulations for the Base Case and the Group Effects Case were run with z_1 distributed as in the Underspecificed Case, the results would remain qualitatively the same, although with smaller biases for Deaton's and the Local Cohorts estimators.

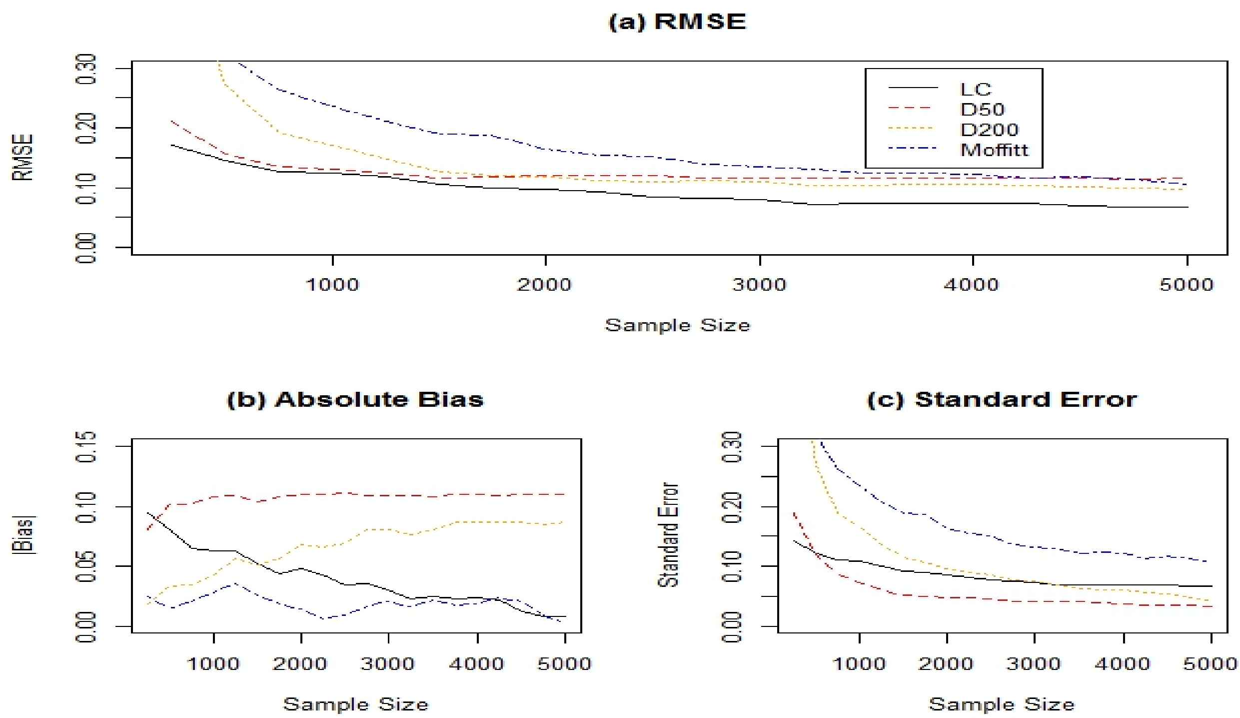
[18] However, because of increased unexplained variability, the Local Cohorts estimator will have a higher variance and may need more observations to sufficiently reduce the small sample bias.

Figure 1: Distribution of LC for Different Sample Sizes: Base Case Simulations



The data generating process in the Base Case satisfies all the assumptions in the Local Cohorts estimator. The vertical line at 2 indicates the true value of the parameters β_0.

Figure 2: Base Case Simulations



The data generating process in the Base Case satisfies all the assumptions in the Local Cohorts estimator. For reference, the true value of the parameter of interest is $\beta_0 = 2$. RMSE, Absolute Bias, and Standard errors are in units (not as percentages of β_0). LC refers to the Local Cohorts estimator; D50 and D200 refer to Deaton's estimator with cohort sizes of approximately 50 and 200 individuals, respectively; and Moffitt refers to Moffitt's estimator implemented with a flexible polynomial to account for the unknown data generating process' functional form.

Base Case. As can be seen from panel (a), the Local Cohorts estimator has a smaller RMSE than the alternative estimators for all sample sizes. Deaton's estimator with cohort size 50 is the next best one in terms of RMSE in the smaller samples, but it is defeated by Deaton's estimator with cohort size 200 for moderate to larger samples. Moffitt's IV estimator is the one with the largest RMSE, except for the larger sample sizes where it is better than Deaton's with cohort size 50.

In terms of the bias, shown in panel (b), Moffitt's IV estimator is clearly superior to all others in small and moderate samples, and performs similarly well to the Local Cohorts estimator for the larger samples.[19] Both versions of Deaton's estimator have rather large biases, and these biases do not decrease in time. The reason for this poor performance in terms of bias is party due to the inflexibility of Deaton's approach, that requires either the fixed-effects for individuals who are very different in terms of time-invariant characteristics to be the similar (the partition of $\mathcal{Z}$ is too coarse) or, when it avoids that, it uses too few observations in each cohort that cannot ensure that the estimates of the fixed-effect for a given cohort is stable across time (the partition of $\mathcal{Z}$ is too fine-graded).

Panel (c) of Figure 2 shows the standard error of each of the estimators for the different sample sizes. Moffitt's IV estimator has the largest standard error of all the estimators considered, which explains why, despite having the smallest bias, it performs poorly in terms of RMSE. The Local Cohorts estimator is somewhere in between the two versions of Deaton's estimator, and has a larger standard error than them for the larger samples. To get a better idea of what these standard errors mean, it is useful to compare them to the standard error from the estimates of an actual panel. For a sample size of 5,000 observations per time period, the Local Cohorts estimator has a standard error that is 32% larger than that of the actual panel (with the same data generating process), whereas Moffitt's IV estimator has a standard error that is 108% larger than the panel's.

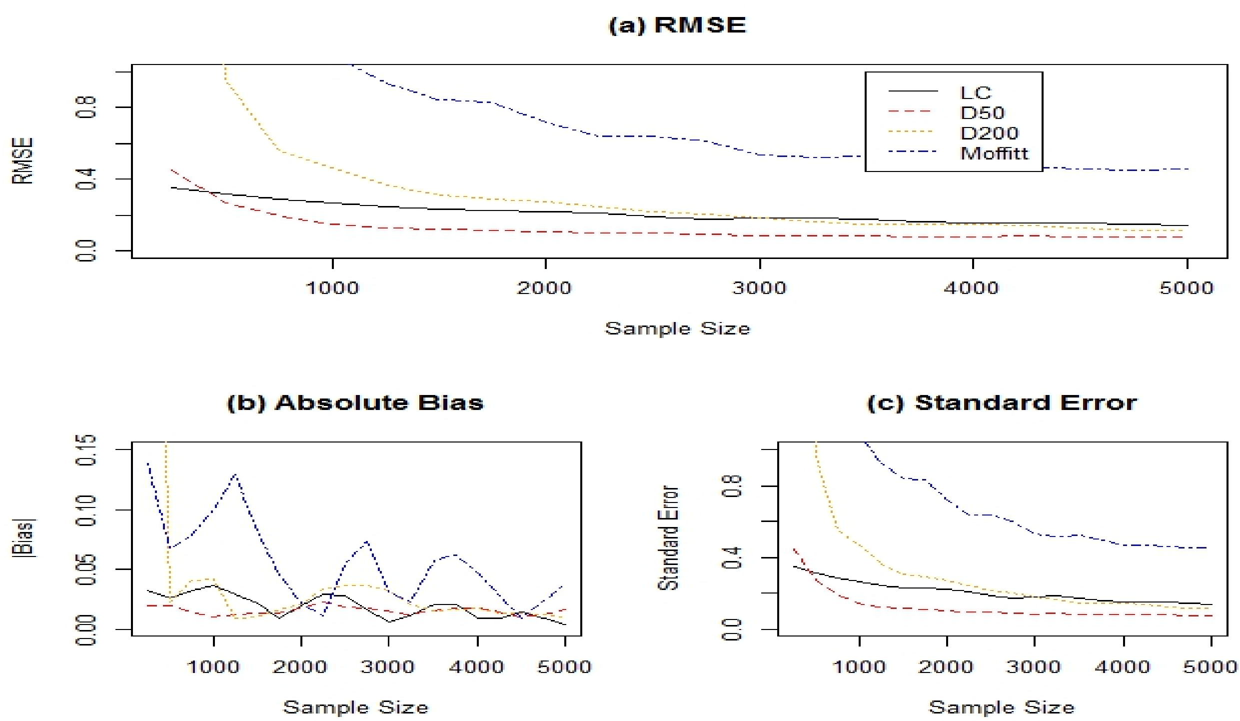
Overall, the Local Cohorts estimator performs better than the other estimators considered. Although Deaton's estimators can have smaller variances, they are biased; and while Moffitt's IV estimator is the one with the smallest bias, it has a much larger variance.

Figure 3 (and Table B2 in the Appendix) show the Monte Carlo results for the Group Effects Case, which satisfies the assumptions in Deaton's approach but violates the continuity assumption (A2) of the local cohorts estimator at 16 distinct points.[20] As expected, Deaton's estimator with cohort size 50 outperform the other estimators in terms of RMSE.

[19]The large variability in the bias of Moffitt's IV estimator is due to the rather small number of Monte Carlo simulations coupled with the large standard error of this estimator. If the Monte Carlo simulations included a significantly larger number of simulations, this bias should be a flat line, approximately at zero. The use of a relatively small number of simulations is due to the computational demands of the LC estimator.

[20]These points are defined by combinations of the first to fourth quintiles of z_1 and z_2.

Figure 3: Group Effects Case Simulations



The data generating process in the Group Effect case assumes that the unobserved individual-level heterogeneity is constant (up to random noise) within groups defined by the quintiles of the observable time-invariant characteristics (which implies a mild violation of assumption A2 of the Local Cohorts estimator). For reference, the true value of the parameter of interest is $\beta_0 = 2$. RMSE, Absolute Bias, and Standard errors are in units (not as percentages of β_0. LC refers to the Local Cohorts estimator; D50 and D200 refer to Deaton's estimator with cohort sizes of approximately 50 and 200 individuals, respectively; and Moffitt refers to Moffitt's estimator implemented with a flexible polynomial to account for the unknown data generating process' functional form.

With respect to the bias, both the Local Cohort estimator and Deaton's perform similarly well, with very small biases. Moffitt's estimator has the highest bias of all estimators in this case. This bias is due to the fact that the flexible polynomial used in the first step of this estimator is not adequate to deal with the discontinuities present in the relation between the unobserved individual-level heterogeneity (f) and the observed individual-level characteristics (z). Panel (c) shows that it is in terms of the standard error that Deaton's estimators outperform the Local Cohorts estimator. The main reason for the better performance of Deaton's estimator in this case is that the relation between f and z follows a step function. This significantly helps Deaton's approach, since it assumes that the fixed-effects are actually group effects, which matches exactly the data generating process in this case. Contrary to this, the Local Cohorts estimator does not assume that the fixed-effects follow a step function, meaning that it loses efficiency in figuring this out.[21] That is, in this case, the more parsimonious nature of Deaton's approach allows for a more precise estimation, with standard errors of about half the size of the Local Cohorts estimator.

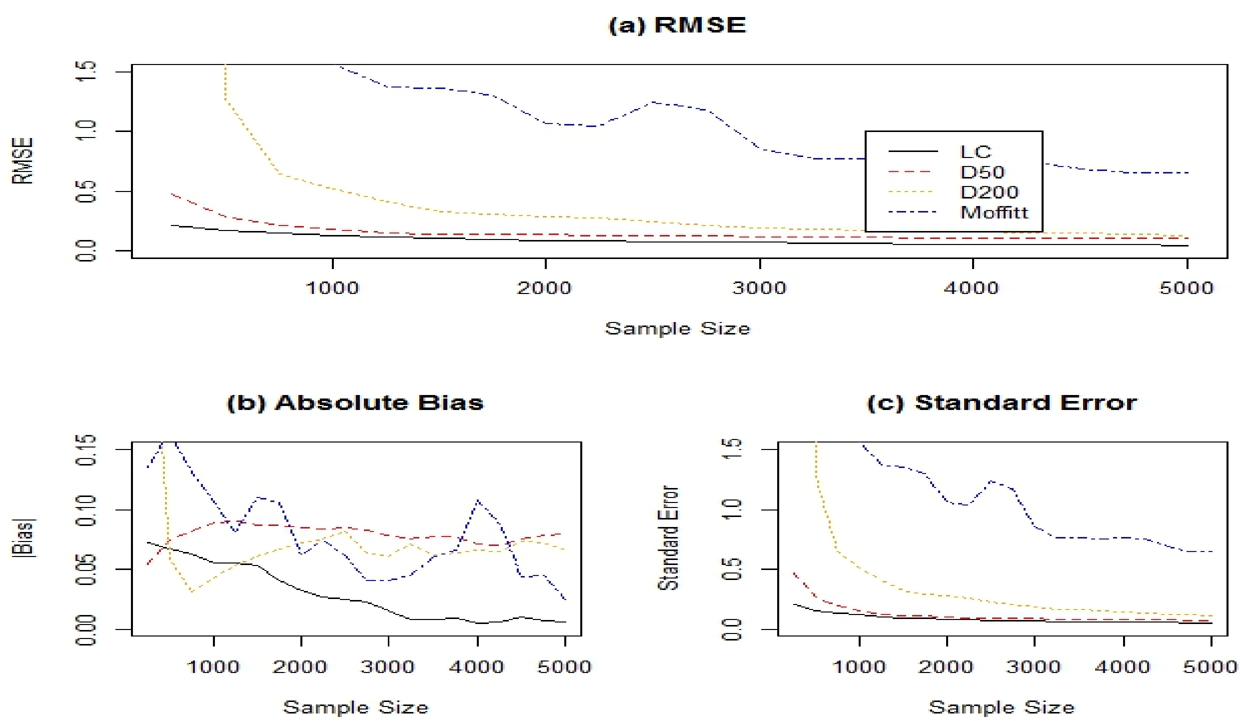
Figure 4 (and Table B3 in the Appendix) shows the Monte Carlo simulation results for the Underspecified Case, in which the fixed-effects depend on three time-invariant characteristics that are correlated with one another, but only two of them are Observable. As can be seen from the simulations, the performance of Deaton's estimators and the LC estiamtor are qualitatively similar to their performance for the Base Case.[22]

The most significant change between the Base Case and the Underspecified Case is the performance of Moffitt's IV estimator. As Panel (b) of Figure 4 shows, this estimator is no longer unbiased. The source of its bias is the misspecification in the first step of the estimator, derived from the unobservability of z_3. This implies that in the second step, the predicted values of x remain correlated with the unobserved heterogeneity from z_3. Both the LC and Deaton's estimator do not suffer significantly from this problem, as in both, the first step estimation

[21]This is similar to comparing the OLS estimates of a linear regression model with nonparametric estimation of the same model.

[22]There are differences in the sizes of the bias and variance term, which are in part due to the different distribution used for z_1.

Figure 4: Underspecified Case Simulations



The data generating process in the Underspecified Case assumes that the unobserved individual-level heterogeneity depends on three time-invariant characteristics (and noise), one of which is unobserved. For reference, the true value of the parameter of interest is $\beta_0 = 2$. RMSE, Absolute Bias, and Standard errors are in units (not as percentages of β_0. LC refers to the Local Cohorts estimator; D50 and D200 refer to Deaton's estimator with cohort sizes of approximately 50 and 200 individuals, respectively; and Moffitt refers to Moffitt's estimator implemented with a flexible polynomial to account for the unknown data generating process' functional form.

integrates over the unobserved time-invariant characteristics.[23]

To summarize, the Local Cohorts estimator has some bias in the smaller samples, but this bias disappears in larger samples, even when the identification assumption (A2) is mildly violated or when there is an unobserved systematic component among the time-invariant characteristics. Deaton's estimator performs better than the Local Cohorts estimator only when the fixed-effects are indeed group effects. However, when this is not the case, it becomes biased, even for the larger sample sizes. Moffitt's estimator generally has no bias, except when its first step is misspecified. This estimator is also very inefficient, always resulting in the highest variance of all the estimators analyzed. Taken together, the Local Cohorts estimator typically outperforms the alternative estimators analyzed here.

5 Comparison to Real Panel Data

In this section I compare the performance of the Local Cohorts estimator developed in this paper with fixed-effects estimates from actual pane data. To do this, I use the National Annenberg Election Study (NAES), 2008 Online Panel. The NAES panel is composed of five waves of interviews conducted over the curse of 16 months, icluding the pre-primary season of the 2008 Presidential Campaign, early primaries, late primaries and party conventions, general election campaign, and post election period. I limit the analysis to the 10,742 respondents that participated in all five NAES waves. I also limit the analysis to the last three waves as these are the ones that include all the variables necessary for the analysis.

The model examined in this section studies the effect of partisan news exposure on respondents' favorability ratings for the Democratic Presidential Nominee, Barack Obama. Favorability of Obama is measured via a feeing thermometer on a scale from 1 (unfavorable views) to 100 (favorable views). To measure exposure to partisan news on television, I use a question that asks respondents to report which programs they watch regularly on television, out of multiple

[23]This integration over unobserved time-invariant characteristics helps reduce the bias from misspecification, but it increases the variance of the estimates.

lists adding up to the 45 most frequently watched news programs, according to Nielsen Ratings. The ideological bias of the 45 TV programs is taken from Dilliplane (2014), who classifies them as liberal, conservative, and neutral.[24] I use several time-invariant characteristics to obtain the LC estimates. These characteristics include age, education, income, party ID, and race, and were obtained from a panel profile wave in NAES. Formally, the model of interest is:[25]

$$\text{ObamaTherm}_{it} = \beta_1 \#\text{LiberalTV}_{it} + \beta_2 \#\text{ConservativeTV}_{it} + \beta_3 \#\text{NeutralTV}_{it} + f_i + \varepsilon_{it} \quad (15)$$

To study the performance of the LC estimator, I generate a sample of repeated cross-sections from the NAES panel in the following way: (1) I draw a random sample of 3,000 respondents from the first wave; (2) draw a random sample of 3,000 respondents from the second wave from among those respondents not sampled in the first wave; (3) draw a random sample of 3,000 respondents from the third wave from among those not sampled in the first or second wave. This ensures that none of the respondents in each of the cross-sections is included in the other ones, so that the resulting dataset resembles real RCS data.

The first, non-parametric, step of the LC estimator requires choosing a kernel. For this application I use a product kernel composed of: (1) the identity function for party ID and race; (2) Gaussian 4th order kernels for age, education, and income. The bandwidths for age, education, and income were selected via cross-validation on the feeling thermometer for Obama.

I compare the results from the LC estimator with those of Moffitt's IV approach, where I use the same time-invariant characteristics plus time dummies in the first stage equation, and to a linear regression that includes the time-invariant characteristics as controls. As ground truth, I use the estimates from a subpanel defined by the individuals in the first repeated cross-section described above.[26]

[24]Table C1 shows the classification of TV programs.

[25]Note that the model specification used in this paper is different from that used in Dilliplane (2014). That paper estimates a fixed effects model using as explanatory variables the average exposure to each type of media for each individual throughout the panel interacted with the panel waves. Thus, the model estimates by Dilliplane (2014) is more a study of individual trends, than it is a traditional fixed-effects analysis.

[26]I do not include a version of Deaton's estimator since this estimator can be though of as a special case of the LC estimator, but where cohorts are arbitrarily defined by the researcher.

Figure 5 shows the estimated coefficients for the effect of the number of liberal, conservative, and neutral TV shows regularly watched by respondents on their favorability ratings of Barack Obama estimated with the actual panel, the LC estimator, and a simple OLS regression with controls. The estimates from Moffitt's IV approach are very large and off target, and are therefore reported separately in Table C2 in the appendix.[27] As the figure shows, the LC estimator generally overestimates the effect of watching all three types of TV programs on the favorability ratings of Obama. While the fixed-effects estimates using the actual panel predict that regularly watching an extra liberal TV show leads to an increase of 0.57 in Obama's favorability ratings, the LC estimates predict an increase of 0.87 points.[28] However, the panel and LC confidence intervals have sufficient overlap so that both estimates are statistically indistinguishable from one another in this case. The estimate from the linear regression with controls, which ignores the unobserved heterogeneity, predicts an increase of 2.1 points in Obama's favorability rating, which is almost 4 times the estimate from the panel. Moffitt's IV estimator instead predicts and increase of 26.9 points.

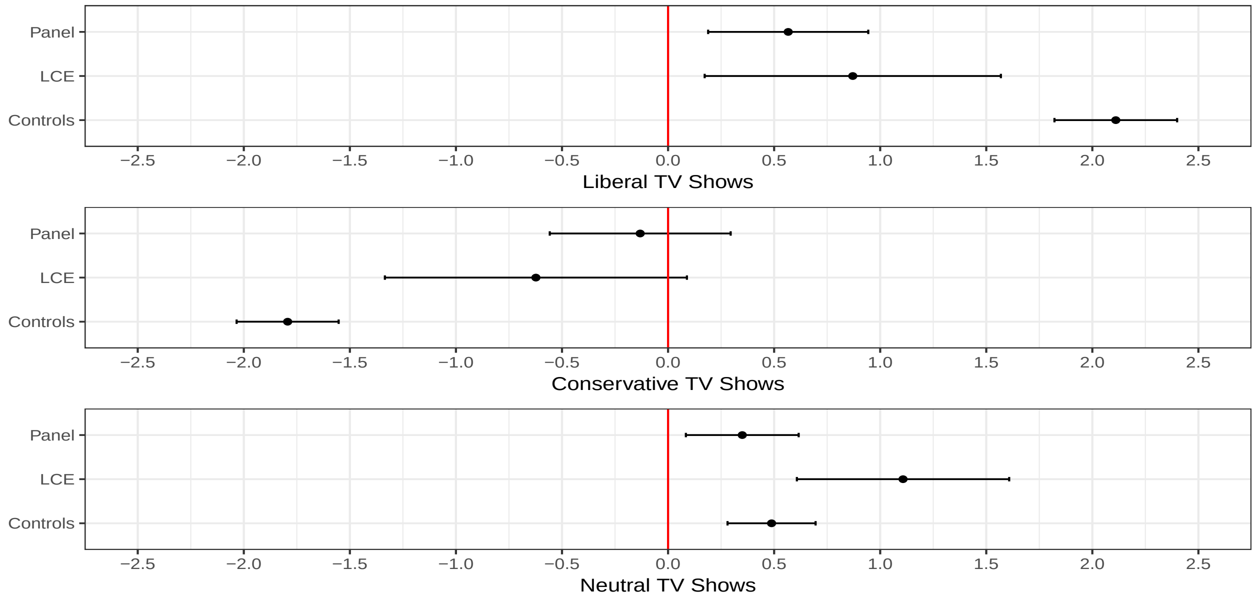
The panel estimates for the number of conservative TV programs watched shows a non-significant effect of -0.13 points, while the LC estimator shows a larger, but still non-significant estimate of -0.6 points. The linear regression with controls and Moffitt's IV estimator, on the other hand, predict a decrease in Obama's rating of 1.8 and 11.6 points, respectively, and are statistically significant.

Finally, the panel fixed-effects estimator predicts an increase of 0.35 points in Obama's rating from watching an extra neutral TV program, while the LC estimator significantly overestimates this quantity, at 1.1 points. The linear regression with controls provides an estimate that is closer to the panel estimate, at 0.49 points. Moffitt's IV estimator is, as with the other variables, significantly misestimating the effect, at -5.7 points.

[27]The most likely reason why Moffitt's IV approach estimates are off target by such a large margin is that the number of TV shows regularly watched in each category resembles an exponential distribution with a small rate. This hampers the ability of the first stage estimates to generate good predictions of the independent variables.

[28]Note that these quantities are actually quite small. Given Obama's average rating of 51.6, it implies and increase of 1.1% or 1.68% according to the panel and LC estimates, respectively.

Figure 5: Coefficient Estimates from an Actual Panel, LCE, and Linear Regression with Controls



Panel estimates are fixed effects estimates from an actual panel of size 3000, based on the same cross-section at time 1 as the LCE and Controls estimates. Controls estimates refer to a linear model of the pooled cross-sections, including the time-invariant characteristics as controls. The variance of the LC estimator was obtained via bootstrap.

Overall, the results for the LC estimator are encouraging. While the LC estimator overestimates the effects of different TV media slants on feeling thermometers for Barack Obama relative to the estimates from an actual panel, this overestimation is not severe. Importantly, the estimates obtained with LC have smaller bias from the ground truth than those obtained with alternative estimators that do not rely on actual panel data. This indicates that the identification assumptions of the LC estimator plausibly hold in this case, notwithstanding the relatively small bias encountered.

6 Discussion

In this paper I develop a semi-parametric two-step estimator procedure, the Local Cohorts estimator, for estimating linear models with individual-level unobserved heterogeneity using repeated cross-sections. I provide identification conditions and derive and asymptotic properties of the estimator, establishing its root-n consistency and asymptotic normality. The identification conditions require that similar individuals have, on average, similar individual-level unobserved heterogeneity. This assumption is typically weaker than the ones used by other estimators, which require that the unobserved individual-level heterogeneity be in fact a group effect with no systematic variation within each group.[29] Strictly speaking, the identification assumptions for the Local Cohorts estimator do not include group effects as a special case, although when these groups are known it can be easily modified to account for them. However, even under unknown groups, this restriction does not represent a significant shortcoming for the LC estimator relative to others, as estimators that rely on group effects have severe problems if the groups are unknown.[30]

In Monte Carlo simulations, I show that the Local Cohorts estimator performs well, with some bias in small samples that disappears in moderate sample sizes, even when the unobserved individual-level heterogeneity comes in the form of group effects. Furthermore, compared to other available estimators, it typically has a smaller RMSE, being only larger than models that assume the presence of group effects when this is indeed the case.

Beyond the Monte Carlo simulations, I compare the Local Cohorts estimator to a fixed-effects estimator derived from actual panel data. The model being estimated seeks to determine the effects of partisan news exposure on the favorability ratings of then presidential candidate Barack Obama. I find that the LC estimator typically outperforms alternative models for RCS data. Moreover, while there is some amount of bias relative to the actual panel estimates, the results show that the identification assumptions of the LC estimator plausibly hold in this case.

Moving forward, there are several avenues for future research on this topic. First, kernel estimators as the ones used in the first step estimation are well-known to be biased in small samples. While the optimal bandwidths for kernel estimators have been derived, it is not clear whether these bandwidths are also optimal for the Local Cohorts estimator, as the relevant

[29] See, for example, Deaton (1985); Inoue (2008).

[30] Deaton (1985) implicitly assumes that there are group effects, but does not establish what these groups, how many there are, or how to discover them from the data.

bias-variance trade-off is the one in the second step estimation, not the first. Second, dimension reduction techniques for the first step estimation can help deal with larger dimensions in the observed individual-level heterogeneity, which can significantly increase the precision of the estimates and help deal with the curse of dimensionality. Third, an extension to discrete outcome models can make this approach more useful to political science applications, where discrete outcomes are common. The asymptotic analysis used in this paper will not suffice for the discrete case, as these models will surely need a larger number of cross-sections because of the incidental parameters problem.

It is possible that the Local Cohorts estimator can prove a useful alternative to the fixed-effects estimators if the efficiency loses are not too large (i.e. when the RCS data is large). This is because RCS data will generally not suffer from attrition as panel data does, and it is cheaper to collect, allowing for significantly larger sample sizes that could more than compensate the efficiency loses relative to the fixed-effects estimator for panel data. Finally, the Local Cohorts could be developed into a complement of fixed-effects panel estimators to help deal with attrition and non-response at certain waves of a panel, as well as boosting the estimates by including in the estimation individuals sampled in cross-sections complementary to the panel design.

References

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, 11:560–596.

Banks, J., Blundell, R., and Preston, I. (1994). Life cycle expenditure allocations and the consumption of costs of children. *European Economic Review*, 76:598–606.

Blundell, R., Browning, M., and Meghir, C. (1994a). Consumer demand and the life-cycle allocation of household expenditures. *Review of Economic Studies*, 61:57–80.

Blundell, R., Duncan, A., and Meghir, C. (1994b). Consumer demand and the life cycle allocation of household expenditures. *Review of Economic Studies*, 58:277–297.

Browning, M., Deaton, A., and Irish, M. (1985). A profitable approach to labor supply and commodity demands over the life cycle. *Econometrics*, 53:503–543.

Collado, M. D. (1997). Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics*, 82:37–67.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30:109–126.

Devereux, P. (2007). Small sample bias in synthetic cohort models of labor supply. *Journal of Applied Econometrics*, 22:839–848.

Dilliplane, S. (2014). Activation, conversion, or reinforcement? the impact of partisan news exposure on vote choice. *American Journal of Political Science*, 58:79–94.

Escanciano, J. C., Jacho-Chávez, D. T., and Lewbel, A. (2012). Identification and estimation of semiparametric two step models. Mimeo.

Escanciano, J. C., Jacho-Chávez, D. T., and Lewbel, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics*, 178:426–433.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag.

Granovsky, B. L. and Muller, H. G. (1991). Optimizing kernel methods: A unifying variational principle. *International Statistical Review*, 59:373–388.

Hahn, J. and Ridder, G. (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81:315–340.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24:726–748.

Ichimura, H. and Lee, S. (2010). Characterization of the asymptotic distribution of semiparametric m-estimators. *Journal of Econometrics*, 159:252–266.

Inoue, A. (2008). Efficient estimation and inference in linear pseudo-panel data models. *Journal of Econometrics*, 142:449–466.

Mammen, E., Rothe, C., and Schienle, M. (2012). Nonparametric regression with nonparametrically generated covariates. *Annals of Statistics*, 40:1132–1170.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared errors. *Annals of Statistics*, 20:712–736.

McKenzie, D. J. (2004). Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics*, 120:235–262.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59:99–123.

Muller, H. G. (1984). Smooth optimum kernel estimator of densities, regression curves and modes. *Annals of Statistics*, 12:766–774.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10:233–253.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics, Volume 4*, pages 2111–3155.

Pelser, B., Eisinga, R., and Franses, P. H. (2002). Inferring transition probabilities from repeated cross sections. *Political Analysis*, 10:113–133.

Pelser, B., Eisinga, R., and Franses, P. H. (2004). Ecological panel inference from repeated cross sections. In King, G., Tanner, M. A., and Rosen, O., editors, *Ecological Inference: New Methodological Strategies*. Cambridge University Press, New York, NY.

Propper, C., Rees, H., and Green, K. (2001). The demand for private medical insurance in the uk: A cohort analysis. *The Economic Journal*, 111:C180–C200.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics, Volume 6B*. Elsevier Science, North-Holland.

Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. In Sevestre, P., editor, *The Econometrics of Panel Data*. Springer-Verlag, Berlin.

Verbeek, M. and Nijman, T. E. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics*, 17:9–23.

Verbeek, M. and Nijman, T. E. (1993). Minimum mse estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, 59:125–136.

Verbeek, M. and Vella, F. (2005). Estimating dynamic models from repeated cross-sections. *Journal of Econometrics*, 127:83–102.

Wand, M. P. and Schucany, W. R. (1990). Gaussian-based kernels. *Canadian Journal of Statistics*, 18:197–204.

Appendix to
Local Cohorts Estimator for Synthetic Panels with Fixed Effects

A Proofs

A.1 Proof of Proposition 1 (Uniform Convergence)

First, notice that by the Triangle Inequality $\sup_{\beta \in \mathcal{B}} \left| G_n(\beta, \widetilde{Y}, \widetilde{X}) - E(g(\beta, \widetilde{Y}_0, \widetilde{X}_0)) \right| \leq$

$$\sup_{\beta \in \mathcal{B}} \left| G_n(\beta, \widetilde{Y}_0, \widetilde{X}_0) - E(g(\beta, \widetilde{Y}_0, \widetilde{X}_0)) \right| + \sup_{\beta \in \mathcal{B}} \left| (nT)^{-1} \sum \sum \left[g(\beta, \widetilde{Y}, \widetilde{X}) - g(\beta, \widetilde{Y}_0, \widetilde{X}_0) \right] \right|$$

Notice that the first term does not depend on the nonparametric first step estimator (but on its true functional form). Therefore, the first term is simply an M-estimator. Since $\mathcal{B}$ is assumed to be compact, the function $g(\cdot)$ is continuous in β, and $E(\sup_{\beta \in \mathcal{B}} g(\beta, \widetilde{Y}_0, \widetilde{X}_0)) < \infty$, this term is $o_p(1)$ by standard results of the Uniform Law of Large Numbers for M-estimators (Newey and McFadden, 1994).

Next, notice that the function $g(\cdot)$ is continuously differentiable (since it is simply a square function). Therefore, it is Lipschitz continuous. This means that $\exists\ \kappa$ such that $\left| g(\beta, \widetilde{Y}, \widetilde{X}) - g(\beta, \widetilde{Y}_0, \widetilde{X}_0) \right| \leq \kappa ||(\widetilde{Y}, \widetilde{X}) - (\widetilde{Y}_0, \widetilde{X}_0)||$, and this holds true for all z (remember that $\widetilde{Y}$ and $\widetilde{X}$ are functions of z). Since this holds for all z, it also holds for the supremum over z. Thus:

$$\left| (nT)^{-1} \sum \sum \left[g(\beta, \widetilde{Y}, \widetilde{X}) - g(\beta, \widetilde{Y}_0, \widetilde{X}_0) \right] \right| \leq \kappa \sup_{z \in \mathcal{Z}} ||(\widetilde{Y}, \widetilde{X}) - (\widetilde{Y}_0, \widetilde{X}_0)||$$

Now, notice that

$$\sup_{z \in \mathcal{Z}} ||(\widetilde{Y}, \widetilde{X}) - (\widetilde{Y}_0, \widetilde{X}_0)|| \leq \sup_{z \in \mathcal{Z}} |\widetilde{Y} - \widetilde{Y}_0| + \sup_{z \in \mathcal{Z}} |\widetilde{x}^1 - \widetilde{x}_0^1| + ... + \sup_{z \in \mathcal{Z}} |\widetilde{x}^k - \widetilde{x}_0^k|$$

where the superscripts in $\widetilde{x}^m$ indicate the mth column of the matrix $\widetilde{X}$

Now, each of the terms in the right hand side of the last inequality can be bounded by:

$$\sup_{z \in \mathcal{Z}} |\bar{r}_{ct} - \bar{r}_{0ct}| + \frac{1}{T} \sup_{z \in \mathcal{Z}} |\bar{r}_{ct} - \bar{r}_{0cs}|$$

where r here stands for y and each of the k dimensions of x. Under assumptions **A3-A5**, each of the terms in the last equation is of order $O_p\left(\left(\frac{\ln(n)}{nh^d}\right)^{1/2} + h^2\right)$, which under appropriate conditions for h is $o_p(1)$ (Hansen, 2008).[31] For example, taking h to be the optimal bandwidth for the Naradaya-Watson estimator (the one that minimizes the mean integrated squared error), $h \propto n^{-1/(4+d)}$, is sufficient for obtaining the $o_p(1)$ rate. Then, given that T is fixed, these are finite sums of $o_p(1)$ terms, and therefore:

$$\sup_{z \in \mathcal{Z}} ||(\widetilde{Y}, \widetilde{X}) - (\widetilde{Y}_0, \widetilde{X}_0)|| = o_p(1)$$

So, putting all of it together, we have that:

$$\sup_{\beta \in \mathcal{B}} |G_n(\beta, \widetilde{Y}, \widetilde{X}) - E(g(\beta, \widetilde{Y}_0, \widetilde{X}_0))| \xrightarrow{p} 0$$

A.2 Proof of Proposition 2 (Consistency)

The consistency is derived from the consistency of M-estimators (see, for example, Theorem 2.1 in Newey and McFadden, 1994). The identification assumption A2, plus the exogeneity assumption A1, ensure that the function $g(\beta, \widetilde{Y}_0, \widetilde{X}_0)$ has a well-separated maximum at β_0. The uniform convergence result from Proposition 1 fulfills the other requirement for consistency.

[31] Newey (1994a,1994b) also provide similar results but under stronger conditions that require z to have bounded support).

A.3 Proof of Proposition 3 (Asymptotic Normality)

The assumptions in Theorem 3.2 in Ichimura and Lee (2010) can be verified, from which the asymptotic normality follows.

- **Asumption 3.1** This assumption requires identification and consistency of the estimator. It is satisfied by the identification restrictions and Proposition 2 (Consistency).

- **Asumption 3.2** This assumption requires the existence a linear approximation of the objective function with a bounded error. It is satisfied by the smoothness of $g(\cdot)$.

- **Asumption 3.3** This assumption requires that a second-order Taylor expansion of $E(g(\cdot))$ be well defined. This assumption is satisfied, again, by the smoothness of $g(\cdot)$ and the continuity assumptions in A5.

- **Asumption 3.4** This assumption imposes a series of smoothness conditions on the first step estimation. Condition (a) in this assumption is satisfied from the identification assumption. Condition (b) is satisfied by the smoothness of the kernels. Condition (c) is satisfied from the uniform convergence results for Kernel estimators (Hansen, 2008, Newey, 1994a and 1994b). Conditions (d) and (e) are satisfied as the first stage estimation does not depend on β_0.

- **Asumption 3.5** This assumption ensures that the remainder term of the Taylor Series expansion is negligible. It can be verified by applying proposition 3.1 of the same paper. Conditions (a) and (b) are satisfied by choosing

$$\omega(\cdot) = \begin{bmatrix} 2 & -2\beta \\ -2\beta & 2\beta^2 \end{bmatrix}$$

  and condition (c) is satisfied as the first step estimation does not depend on β_0.

- **Asumption 3.6** This assumption requires that the effect of the first stage estimation on the final precision of the estimates of β_0 be representable as the sum of zero-mean and finite-variance random variables. It can be verified by using the results from remark 3.3 of the same paper, and defining:

$$g(z, \theta) = \begin{bmatrix} E(2\tilde{y} - \beta\widetilde{x}) \\ E(-2\beta(\widetilde{y} - \beta\widetilde{x})) \end{bmatrix}$$

and

$$E\big(\varphi(z, \theta)|\nu(Z) = \nu(z)\big) = \begin{bmatrix} E(y|Z = z) \\ E(x|Z = z) \end{bmatrix}$$

B Tables from Simulations

Table B1: Bias, Standard Error, and Root Mean Square Error for Simulations of the Base Case

Sample Size	RMSE				Absolute Bias				Standard Error			
	LC	D50	D200	M	LC	D50	D200	M	LC	D50	D200	M
250	0.17	0.21	0.54	0.43	0.10	0.08	0.02	0.02	0.13	0.19	0.53	0.43
500	0.15	0.16	0.27	0.32	0.08	0.10	0.03	0.01	0.12	0.12	0.27	0.32
750	0.13	0.13	0.19	0.26	0.06	0.10	0.03	0.02	0.11	0.09	0.19	0.26
1000	0.12	0.13	0.17	0.24	0.06	0.11	0.04	0.03	0.11	0.07	0.16	0.23
1250	0.12	0.12	0.15	0.21	0.06	0.11	0.06	0.04	0.10	0.06	0.14	0.21
1500	0.11	0.12	0.13	0.19	0.05	0.10	0.05	0.03	0.09	0.05	0.12	0.19
1750	0.10	0.12	0.12	0.19	0.04	0.11	0.06	0.02	0.09	0.05	0.11	0.19
2000	0.10	0.12	0.12	0.17	0.05	0.11	0.07	0.01	0.08	0.05	0.10	0.16
2250	0.09	0.12	0.11	0.15	0.04	0.11	0.07	0.01	0.08	0.05	0.09	0.15
2500	0.08	0.12	0.11	0.15	0.03	0.11	0.07	0.01	0.08	0.05	0.08	0.15
2750	0.08	0.12	0.11	0.14	0.04	0.11	0.08	0.02	0.07	0.04	0.08	0.14
3000	0.08	0.12	0.11	0.13	0.03	0.11	0.08	0.02	0.07	0.04	0.08	0.13
3250	0.07	0.12	0.10	0.13	0.02	0.11	0.08	0.02	0.07	0.04	0.07	0.13
3500	0.07	0.12	0.10	0.12	0.02	0.11	0.08	0.02	0.07	0.04	0.06	0.12
3750	0.07	0.12	0.11	0.12	0.02	0.11	0.09	0.02	0.07	0.04	0.06	0.12
4000	0.07	0.12	0.11	0.12	0.02	0.11	0.09	0.02	0.07	0.04	0.06	0.12
4250	0.07	0.11	0.10	0.12	0.02	0.11	0.09	0.02	0.07	0.04	0.06	0.11
4500	0.07	0.12	0.10	0.12	0.01	0.11	0.09	0.02	0.07	0.04	0.05	0.12
4750	0.07	0.11	0.10	0.11	0.01	0.11	0.09	0.01	0.07	0.03	0.05	0.11
5000	0.07	0.11	0.10	0.11	0.01	0.11	0.09	0.00	0.07	0.03	0.04	0.11

The data generating process in the Base Case satisfies all the assumptions in the Local Cohorts estimator. LC refers to the Local Cohorts estimator, D50 and D200 refer to Deaton's estimator with cohort sizes of 50 and 200 individuals, and M refers to Moffitt's estimator using a flexible polynomial to account for the unknown data generating process. For reference, the true value of the parameter of interest is $\beta_0 = 2$.

Table B2: Bias, Standard Error, and Root Mean Square Error for Simulations of the Group Effects Case

Sample Size	RMSE				Absolute Bias				Standard Error			
	LC	D50	D200	M	LC	D50	D200	M	LC	D50	D200	M
250	0.35	0.45	4.42	1.83	0.03	0.02	0.68	0.14	0.35	0.45	4.36	1.82
500	0.31	0.27	0.96	1.28	0.03	0.02	0.02	0.07	0.31	0.27	0.96	1.27
750	0.29	0.20	0.55	1.15	0.03	0.01	0.04	0.08	0.29	0.20	0.55	1.15
1000	0.27	0.15	0.47	1.09	0.04	0.01	0.04	0.10	0.26	0.15	0.46	1.08
1250	0.24	0.13	0.36	0.93	0.03	0.01	0.01	0.13	0.24	0.12	0.36	0.93
1500	0.23	0.12	0.31	0.84	0.02	0.01	0.01	0.08	0.23	0.12	0.31	0.84
1750	0.23	0.11	0.29	0.83	0.01	0.01	0.02	0.05	0.23	0.11	0.29	0.83
2000	0.22	0.11	0.27	0.72	0.02	0.02	0.02	0.02	0.22	0.10	0.27	0.72
2250	0.21	0.10	0.24	0.64	0.03	0.02	0.03	0.01	0.21	0.10	0.24	0.64
2500	0.19	0.10	0.22	0.64	0.03	0.02	0.04	0.05	0.19	0.09	0.21	0.64
2750	0.17	0.09	0.20	0.61	0.02	0.02	0.04	0.07	0.17	0.09	0.20	0.60
3000	0.18	0.08	0.18	0.53	0.01	0.02	0.03	0.03	0.18	0.08	0.18	0.53
3250	0.19	0.09	0.16	0.52	0.01	0.01	0.02	0.02	0.18	0.09	0.16	0.52
3500	0.17	0.08	0.15	0.52	0.02	0.01	0.01	0.06	0.17	0.08	0.15	0.52
3750	0.16	0.08	0.15	0.50	0.02	0.02	0.02	0.06	0.16	0.08	0.14	0.49
4000	0.15	0.08	0.15	0.47	0.01	0.02	0.02	0.05	0.15	0.08	0.15	0.47
4250	0.15	0.08	0.14	0.47	0.01	0.01	0.01	0.03	0.15	0.08	0.14	0.47
4500	0.15	0.08	0.13	0.46	0.01	0.01	0.01	0.01	0.15	0.08	0.12	0.46
4750	0.15	0.08	0.12	0.45	0.01	0.01	0.01	0.02	0.15	0.07	0.11	0.45
5000	0.14	0.07	0.12	0.45	0.00	0.02	0.01	0.04	0.14	0.07	0.12	0.45

The data generating process in the Group Effects Case assumes that the unobserved individual-level heterogeneity is constant (up to random noise) within groups defined by the quintiles of the observable time-invariant characteristics (which implies a mild violation of assumption A2 of the Local Cohorts estimator). LC refers to the Local Cohorts estimator, D50 and D200 refer to Deaton's estimator with cohort sizes of 50 and 200 individuals, and M refers to Moffitt's estimator using a flexible polynomial to account for the unknown data generating process. For reference, the true value of the parameter of interest is $\beta_0 = 2$.

Table B3: Bias, Standard Error, and Root Mean Square Error for Simulations for the Underspecified Case

Sample Size	RMSE				Absolute Bias				Standard Error			
	LC	D50	D200	M	LC	D50	D200	M	LC	D50	D200	M
250	0.22	0.47	18.07	2.15	0.07	0.05	0.33	0.13	0.20	0.47	18.06	2.14
500	0.17	0.28	1.27	2.00	0.07	0.07	0.06	0.17	0.15	0.27	1.27	1.99
750	0.15	0.21	0.64	1.81	0.06	0.08	0.03	0.13	0.13	0.20	0.64	1.80
1000	0.13	0.18	0.51	1.57	0.06	0.09	0.04	0.11	0.12	0.16	0.51	1.56
1250	0.11	0.15	0.41	1.37	0.06	0.09	0.05	0.08	0.10	0.12	0.40	1.37
1500	0.11	0.14	0.33	1.36	0.05	0.09	0.06	0.11	0.09	0.11	0.32	1.35
1750	0.10	0.14	0.30	1.31	0.04	0.09	0.07	0.11	0.09	0.11	0.29	1.30
2000	0.09	0.13	0.29	1.06	0.03	0.09	0.07	0.06	0.08	0.10	0.28	1.06
2250	0.08	0.13	0.27	1.04	0.03	0.08	0.08	0.07	0.08	0.09	0.26	1.04
2500	0.08	0.12	0.24	1.24	0.03	0.08	0.08	0.06	0.07	0.09	0.23	1.24
2750	0.07	0.12	0.21	1.18	0.02	0.08	0.06	0.04	0.07	0.09	0.20	1.18
3000	0.07	0.12	0.19	0.85	0.01	0.08	0.06	0.04	0.07	0.09	0.18	0.85
3250	0.06	0.11	0.18	0.77	0.01	0.07	0.07	0.05	0.06	0.08	0.17	0.76
3500	0.06	0.11	0.17	0.76	0.01	0.08	0.06	0.06	0.06	0.08	0.16	0.76
3750	0.06	0.11	0.17	0.76	0.01	0.08	0.06	0.07	0.06	0.08	0.15	0.75
4000	0.05	0.10	0.16	0.77	0.01	0.07	0.07	0.11	0.05	0.08	0.14	0.76
4250	0.05	0.10	0.15	0.76	0.01	0.07	0.06	0.09	0.05	0.07	0.13	0.75
4500	0.05	0.11	0.15	0.69	0.01	0.07	0.07	0.04	0.05	0.07	0.13	0.69
4750	0.05	0.11	0.14	0.65	0.01	0.08	0.07	0.04	0.05	0.07	0.12	0.65
5000	0.05	0.11	0.13	0.65	0.01	0.08	0.07	0.02	0.05	0.07	0.11	0.65

The data generating process in the Underspecified Case assumes that the unobserved individual-level heterogeneity depends on three time-invariant characteristics (and noise), one of which is unobserved. LC refers to the Local Cohorts estimator, D50 and D200 refer to Deaton's estimator with cohort sizes of 50 and 200 individuals, and M refers to Moffitt's estimator using a flexible polynomial to account for the unknown data generating process. For reference, the true value of the parameter of interest is $\beta_0 = 2$.

C Tables from Application

Table C1: Classification of TV Programs

Liberal	Conservative	Neutral
Anderson Cooper 360	Fox News	ABC World News
Countdown with Keith Olbermann	Fox Report with Shepard Smith	CBS Evening News
Hardball with Chris Matthews	Hannity and Colmes	Lou Dobbs
CNN Headline News/Newsroom	Hannity's America	NBC Nightly News
ABC News Nightline	The O'Reilly Factor	Meet the Press
Situation Room w. Wolf Blitzer	The Beltway Boys	Today Show
The Daily Show w. Jon Stewart	Studio B w. Shepard Smith	The NewsHour w. Jim Lehrer
Good Morning America	Geraldo at Large	Larry King Live
This Week w. G. Stephanopoulos	Your World with Neil Cavuto	60 Minutes
The View	Fox and Friends	Face the Nation
The Colbert Report	Special Report w. Brit Hume	Reliable Sources
Late Edition w. Wolf Blitzer		The Early Show
MSNBC Live		Frontline
Out in the Open		CBS Sunday Morning
BET News		20/20
		Dateline NBC
		The McLaughin Group
		CBS Morning News
		America This Morning

The classification of TV programs comes from Dilliplane (2014).

Table C2: Estimates from NAES 2008 Data

	Lib TV	L	U	Con TV	L	U	Neu TV	L	U
Panel	0.57	0.19	0.94	-0.13	-0.56	0.30	0.35	0.08	0.62
LCE	0.87	0.17	1.57	-0.62	-1.33	0.09	1.11	0.61	1.61
Controls	2.11	1.82	2.40	-1.79	-2.03	-1.55	0.49	0.28	0.70
Moffitt's	26.93	23.00	30.86	-11.63	-13.19	-10.07	-5.72	-8.30	-3.15

Panel refers to estimates obtained from actual panel data. Controls refers to a linear regression that includes the observed time-invariant characteristics as controls. L and U denote the lower and upper bounds of 95% confidence intervals for the corresponding coefficients.