Imputing Missing Waves for Pseudo Panels: A Generalized Scoring and Matching Method*

Zhongjian Lin† Esfandiar Maasoumi‡
Emory University Emory University

March 29, 2020

Abstract

We propose a scoring method to impute missing cross section waves in a would be panel setting, or counterfactual samples. We identify statistical "matches" in different cross sections based on a multicomponent "optimal aggregator". This is similar to propensity score matching but is applicable even when a classifier outcome (e.g., treatment status) is not observed. It is also substantively different from the cohort methods and other "model dependent" approaches. We obtain and model the actual observed covariates as surrogates, not the cohort averages that satisfy a given model. This is also a distinction between our approach and the "synthetic variable" method, where synthetic variates are used for missing observations. Imputation here is by assignment based on an "aggregate score" which is obtained from any desired group of observed covariates. The aggregation method is "information efficient". Our method is relevant for longitudinal (panel) grouping, as in network memberships, in treatment effect settings, and for missing data. In the case of pseudo panels, we assume traditional unobserved heterogeneity and error component structures for the observed and missing cross section waves. Our method is assessed for simulated panel data, and in two notable empirical studies, one on the private return to R&D in the presence of spillovers using macropanel data, and the other on female labor force participation using micropanel data (PSID) where we know the entire panel.

Keywords: Panel data, Unobserved Heterogeneity, Missing Data, Repeated Cross-Section, Stitching, Information Aggregation, Propensity Score, Treatment Effect, R&D, Female Labor Force Participation.
JEL Classifications: C21, C23, J21, O32.

---

*We are grateful to Stéphane Bonhomme, Ivan Fernández-Val, Kei Hirano, Serena Ng, Martin Weidner and participants at numerous seminars and conferences for useful comments and suggestions.

†Department of Economics, Emory University, 1602 Fishburne Drive, Atlanta, GA 30307. Email: zhongjian.lin@emory.edu.

‡Department of Economics, Emory University, 1602 Fishburne Drive, Atlanta, GA 30307. Email: esfandiar.maasoumi@emory.edu.

1 Introduction

Panel data allow for control of unobserved heterogeneity and facilitate identification and estimation of structural parameters, allowing fixed effects and unobserved heterogeneity. See Chamberlain (1984); Arellano and Honoré (2001); Arellano (2003); Baltagi (2008); Wooldridge (2010); Arellano and Bonhomme (2011); Su and Ullah (2011); Hsiao (2014); Arellano and Bonhomme (2017).

"Micropanels" provide longitudinal observations of the same households or firms from surveys, census, administrative records, or company balance accounts. Typically, micropanels consist of large number of individuals (large N) for short time periods. Short time micropanels are inadequate for longer life-cycle dynamic analyses. Furthermore, attrition can dramatically shrink the cross sectional dimension and adversely impact inferences. Less expensive cross section samples are available over time, such as the Current Population Survey (CPS), but lack the desired repeated observations (longitudinal) benefits of real panels. This is also the case for many financial firms and portfolios that may evolve over time.

In this paper we explore a new approach to creating pseudo panels from cross section samples of randomly selected subjects at different times. Our solution is not model based, in contrast to almost all other imputation methods. Model based imputed "missing" data should not be used to assess model specification and in related inferences. We seek to discover "statistical twins", subjects that are similar, or matches in different cross sections. This is achieved in two steps: first we obtain ranks by aggregating multiple indicators and characteristics of subjects in an information efficient manner (similar to propensity score assessment). In the second step, we stitch individual cross section units over time that have similar ranks. We then adopt familiar error component and other factor structure for these pseudo panels.

There are several approaches that are similar to ours. The similarity is often more in terms of the problem that is addressed, rather than the solution. For example, the popular method of "cohorts" is different in at least two main respects: Firstly, one cohort methods employs a narrow definition of "similarity" between subjects based on a one or two attributes. Secondly, cohort methods switch to modelling cohort averages to explain cohort average outcomes. In an influential paper, Deaton (1985) proposed a "cohort" method for estimating structural parameters based on Family Expenditure Surveys (FES). Also, Browning, Deaton, and Irish (1985) used the cohort analysis to study life-cycle consumption and labor supply using the same time series of cross section source. Membership in a cohort is defined deterministically by age and work type. This presumes a great deal of homogeneity in all other characteristics and dimensions, observed and otherwise. This approach is also a *model based* imputation method for missing data. [1] The cohort method is essentially a *data combination* solution to missing data. Data combination methods complement the target sample with auxiliary data(Hellerstein and Imbens, 1999; Hirano, Imbens, Ridder, and Rubin, 2001; Chen, Hong, and Tarozzi, 2008; Inoue and Solon, 2010; Fan, Sherman, and Shum, 2014; Graham, Pinto, and Egel, 2016). We also adopt a monotonicity condition across periods that is less demanding than the assumption of a common model across samples, favored in data combination methods. For more details, see Ridder and Moffitt (2007). We merely assume the same probability law and a common covariance structure for the observable and unobservables. No particular model specification is imposed on the imputation process. This provides a robustness to model misspecification that is not shared by many approaches to missing data problems. In addition, the cohort and synthetic variate techniques are grouping methods based on averages that may fail to preserve important heterogeneity features within groups and/or nonlinear effects and relations. Furthermore, in instrumental variables (IV) inference based on cohort and similar methods, one requires interaction terms for cohort dummies and time dummies, for validity of instruments for all covariates in the model. This provides further challenges to finding valid instruments. For more details regarding the limitations of cohort analysis, see Verbeek and Vella (2005); Verbeek (2008), and Moffitt (1993).

A seemingly similar method is the synthetic control method proposed in Abadie and

[1] The cohort approach is very common in studies of life-cycle consumption, saving and labor supply, Attanasio and Weber (1993); Blundell, Browning, and Meghir (1994); Attanasio and Weber (1995); Attanasio (1998); Blundell, Duncan, and Meghir (1998); Attanasio, Banks, Meghir, and Weber (1999); Gourinchas and Parker (2002); Fernández-Villaverde and Krueger (2007); Fernandez-Villaverde and Krueger (2011); Pagel (2017) to name a few.

Gardeazabal (2003); Abadie, Diamond, and Hainmueller (2010). A weighted average of observed covariates and outcomes for "untreated" subjects is used as a synthetic counterfactual observation, and modelled in a factor component panel setting with many untreated subjects, in order to estimate an average treatment effect. Optimal estimators for weights for components of the synthetic average are chosen to maximize "similarity" with the observed covariates for the treated (or pretreatment) based on a quadratic/Mahanalobis type distance measure. This notion of similarity is comparable to our information theoretic measure (described shortly), but we do not employ or model synthetic averages which must conform to a given stochastic model for micro level observations. In our proposal, units in successive cross sections retain their actual observed values, much as in propensity score methods.

To motivate our method, we note that extrapolation of *similarity* based on observables is fundamental to learning, as enshrined in treatment effects and program evaluation [2].

In this paper, we propose ideal aggregates/indices of observed covariates for individual units, and assess their closeness to other units, in different waves or data sources, with similar index values. We do not compare cross section units by their outcomes. If and when an additional suitable assignment or outcome is available, as in treatment status, our approach is applicable and takes on the form of an information efficient "propensity score". Our aggregate functions also provide scores, but they may be estimated without logit or probit type regressions. In this paper we will focus on one-to-one matching, but multiple matching is readily accommodated.

The paper unfolds as follows. We begin with a description of general FE models in panel data and identify the missing counterfactuals. Section 3 describes the rule for aggregation of observed covariates in several dimensions and its relation to unobserved individual heterogeneity. Section 4 describes an "information efficient" technique for aggregation over many characteristics. Estimation of the parameters in the optimal aggregator is also discussed. Section 5 provides simulation evidence, and Section 6

[2]Schank (1986); Riesbeck and Schank (1989); Gilboa and Schmeidler (2003). Gilboa and Schmeidler (2003) provide an axiomatic analysis of the inductive inference and study the way that possible predictions are ranked, as a function of past observations.

presents an empirical application to private return to R&D in the presence of spillovers, and an application to female labor force participation. Section 7 concludes.

2 Panel Data Models

Based on a panel of observations, or one constructed by matched individuals based on some aggregation index S_i, from different periods, we would have a (syntetic) panel, $\{Y_{it}, X_{it}\}_{i=1,\cdots,n;t=1,\cdots,T}$. In panel data models it is commonly assumed that the ith individual possesses the same (or similar) unobserved heterogeneity over time. We illustrate with three examples

Example 1 (Linear Fixed Effects Panel Data Model)**.** *Consider the linear fixed effects model:*

$$Y_{it} = \alpha_i + X_{it}^T\beta + \varepsilon_{it}.$$

The least square dummy variable estimator is then

$$\hat{\beta} = \Big[\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it} - \overline{X}_i)(X_{it} - \overline{X}_i)^T\Big]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it} - \overline{X}_i)(Y_{it} - \overline{Y}_i).$$

Example 2 (Additive Nonlinear Regression)**.** *Consider the nonlinear fixed effects model:*

$$Y_{it} = \alpha_i + g_t(X_{it}, \beta) + \varepsilon_{it}.$$

The optimal estimator is the nonlinear within-group estimator:

$$\hat{\beta}_{NWG} = \underset{\beta}{argmin}\sum_{i=1}^{n}\Big[Y_i - g(X_i, \beta)\Big]^T\cdot\Big(I_T - ll^T/T\Big)\Big[Y_i - g(X_i, \beta)\Big],$$

where $Y_i = (Y_{i1}, \cdots, Y_{iT})^T$, $X_i = (X_{i1}, \cdots, X_{iT})^T$, $g(\cdot) = (g_1(\cdot), \cdots, g_T(\cdot))^T$ *and* $l = (1, \cdots, 1)^T$.

Example 3 (Discrete Choice Model)**.** *Consider the binary choice model:*

$$Y_{it} = \boldsymbol{1}\Big\{X_{it}^T\beta + \alpha_i - \varepsilon_{it} > 0\Big\},$$

where ε follows standard Logistic distribution. A consistent estimator for structural parameter, β obtains from conditional MLE (see Andersen, 1973; Cox, 1958; Chamberlain, 1980; Rasch, 1960). There are many bias reduction methods for discrete panel data models, see Hahn and Newey (2004); Arellano and Carrasco (2003); Arellano and Hahn (2007); Greene (2004); Bester and Hansen (2009); Dhaene and Jochmans (2015); Fernández-Val (2009); Fernández-Val and Vella (2011); Fernández-Val and Weidner (2016); Honoré and Lewbel (2002). There are methods to deal with dynamic discrete choice panel data models, see Hahn and Kuersteiner (2011); Honoré and Kyriazidou (2000); Carro (2007).

The requirements for applicability of our pseudo panel construction to random coefficient panel data models is deferred to future research.

3 Analysis of Unobserved Heterogeneity

Accounting for unobserved individual heterogeneity often leads to significant and substantively different inferences. Fixed effects approaches in panel data models are attractive due to their flexible treatment of unobserved heterogeneity.

We extrapolate the ranks of unobserved heterogeneity of individuals from the information embedded in their observed characteristics. For illustration, we consider individuals 1 and 2 in a given period. Denote the unobserved heterogeneity and observed characteristics of individual 1 and 2, respectively, as (α_1, X_1) and (α_2, X_2). We make the following well known assumption on "rank similarity":

Assumption 1. *Observed characteristics ranks represent unobserved heterogeneity ranks, i.e. $\alpha_1 > \alpha_2$ iff $S(X_1) > S(X_2)$ for some appropriate aggregation function $S(\cdot)$*[3].

Remark 1. *Assumption 1 is similar to the axiomatic prediction rule in Gilboa and Schmeidler (2003). In life-cycle consumption and labor supply models, MaCurdy (1981) suggests taking unobservable marginal utility, λ, constant over the lifetime of the consumer, and treat it as an unobserved heterogeneity in the panel data analysis. MaCurdy (1981) also points out that*

[3] We defer the discussion of "appropriate" information aggregation to Section 4. Single variate cohort assignment is a narrow aggregation method.

"it is theoretically possible to compute a unique value for λ using data on an individual's consumption, labor supply and wage rate at a point in time." Thus the observed consumption, labor supply and wage rate provide information for extrapolation/imputation of the unobserved marginal utility constant in the life-cycle model.

Similarly, in the return to schooling studies, (see Card, 1999; Heckman, Lochner, and Todd, 2006, for surveys) ability is usually considered as an unobserved heterogeneity, while observed test scores may indicate the rank in ability. It is clear that one does not have to employ all the observed characteristics in aggregation[4]*. There are contexts in which rank similarity is not compelling.*

Assumption 1 itself does not suffice to make the fixed effects approaches applicable to our pseudo panels. An additional assumption is required, as follows

Assumption 2. *The distribution of unobserved heterogeneity is the same in different periods.*

While this implies the same covariance structure for our pseudo panel as the "true underlying panel", it does not require further knowledge or specification of the distribution in the above assumption. Consider a series of independent cross sections, $\{Y_{i1}, X_{i1}\}_{i=1,\cdots,n}$ and $\{Y_{i'2}, X_{i'2}\}_{i'=1,\cdots,n}$[5]. Denote the unobserved heterogeneity in periods 1 and 2 as $\alpha_i, i = 1, \cdots, n$ and $\alpha_{i'}, i' = 1, \cdots, n$. Assumption 2 states that $F_\alpha(\cdot)$ and $F'_\alpha(\cdot)$ are the same where $F_\alpha(\cdot)$ and $F'_\alpha(\cdot)$ are the distribution functions of α_i and $\alpha_{i'}$, respectively. Waves of cross sections usually contain representative samples of the population. For instance, the Current Population Surveys (CPS) is a monthly representative sample of the United States labor force and many other covariates. Representativeness validates the same-distribution condition in Assumption 2. Combination of Assumptions 1 and 2 makes inference based on the stitched pseudo panel feasible. We summarize the above as a proposition:

Proposition 1. *Given Assumptions 1 and 2, one to one stitching for individuals in different periods is valid. The stitched pairs have similar unobserved heterogeneity.*

[4]This would mitigate the large dimension problem discussed in Section 4. However, "Fundamentalism", suggests that we take as many characteristics as possible for proper "representation"

[5]We set $T = 2$ for illustration. Our results generalize to larger T. Without loss of generality, we assume the numbers of observations along time dimension are the same.

Remark 2. Proposition 1 may be generalized to multiple-to-one stitching, the so called multiple imputation in some contexts. The empirical similarity literature indicates how one may proceed, see Gilboa, Lieberman, and Schmeidler (2006, 2011); Gayer, Lieberman, and Yaffe (2019). The method also applies to the general framework of panel data with factors and interactive effects (e.g. Pesaran, 2006; Bai, 2009; Su and Jin, 2012; Moon and Weidner, 2015; Chen, Fernández-Val, and Weidner, 2019).

A seemingly similar approach is propensity score matching in treatment effects studies, e.g. Rosenbaum and Rubin (1983); Heckman, Ichimura, and Todd (1998); Abadie and Imbens (2006, 2016). Scoring by the above aggregation function $S(\cdot)$ is similar to the propensity score function, but we do not need to run probit or logit type regressions that require an observed outcome (e.g., treatment status). Further, the components of $S(\cdot)$ need not be time-invariant in our approach. The common trend condition is sufficient to preserve the rankings.

We note that in (unbalanced) panels with missing observations, factor-based imputation has been proposed to impute missing observations with estimated factors and their loading parameters (with or without missing at random assumption), (see Bai and Ng, 2019; Su, Miao, and Jin, 2019). Our method can also be used to impute missing observations by extrapolating from matched units.

4 Information Aggregation of Observed Characteristics

In this section, we describe first an optimization approach to construction of ideal aggregates/score functions of several desired characteristics. The optimal aggregation is for each cross section wave (period) and we suppress the subscript t when it is not confusing. In contrast to estimation of propensity scores we do not have to have an observed outcome (treatment assignment). A probit or logit type estimation is not feasible.

Maasoumi (1986) proposed an optimal aggregator, denoted by S, to summarize the entire distribution information from several observed characteristics, such as income, health, and education. The optimal aggregation function minimizes the generalized relative entropy divergence between the aggregator S_i and each of its components X_{ij}. Since not all available characteristics may be used in aggregation, we denote by Z_i the subset of observed characteristics X_i employed for scoring, with dimension d.

The optimal aggregator/score function is formally defined as the minimizer of the following generalized relative entropy criterion that is sometimes referred to as the Cressie-Read family:

$$S(Z_i) = \underset{S}{argmin}\ D_\delta(S, Z; \gamma) = \sum_{j=1}^{d} \gamma_j \left[\sum_{i=1}^{n} S_i \Big[\Big(S_i/Z_{ij} \Big)^{\delta} - 1 \Big] / \delta(\delta+1) \right]\ s.t. \sum_{i=1}^{n} S_i = 1. \quad (1)$$

We have the optimal aggregator as

$$\begin{aligned} & S_i \propto \left(\sum_{j=1}^{d} \gamma_j Z_{ij}^{-\delta} \right)^{-1/\delta}, \quad \delta \neq 0, -1; \\ & S_i \propto \Pi_{j=1}^{d} Z_{ij}^{\gamma_j}, \quad \delta = 0; \\ & S_i \propto \sum_{j=1}^{d} \gamma_j Z_{ij}, \quad \delta = -1. \end{aligned} \quad (2)$$

By minimizing the "divergence measure" D_δ, we make the **vector** $S \equiv (S_1, S_2, \cdots, S_n)$ as close to their corresponding multivariate attributes as possible, see Maasoumi (1986) for details. From Information Theory, the vector S absorbs all the objective statistical information in the data, and any deviation from S will be accordingly suboptimal.

4.1 Estimation of Parameters in the Optimal Aggregator

We require estimates for the unknown parameters in the aggregate score function. There are subjective methods, of course, but we demonstrate two data based methods to derive the weights for different attributes($\gamma_j' s$) and substitution degree between attributes(δ). Hereafter, we focus on Equation (2) with $\delta \neq 0, -1$.

4.1.1 Data Driven Two-Step Estimation

When $d = 2$, we can adopt a two-step procedure proposed by Maasoumi and Racine (2016) to estimate the parameters. The first step is the nonparametric estimation of the conditional pdf, cdf and quantiles of the appropriate covariates. The second step is a standard fitting regression. Explicitly, consider the (CES) aggregator function

$$S(Z_i) = A\left(\gamma Z_{i1}^{-\delta} + (1-\gamma) Z_{i2}^{-\delta}\right)^{-1/\delta}. \tag{3}$$

Taking partial derivatives with respect to Z_{i1} and Z_{i2} results in

$$S_1 \equiv \frac{\partial S(Z_i)}{\partial Z_{i1}} = A\gamma\left(\gamma Z_{i1}^{-\delta} + (1-\gamma) Z_{i2}^{-\delta}\right)^{(-1/\delta)-1} Z_{i1}^{-\delta-1},$$
$$S_2 \equiv \frac{\partial S(Z_i)}{\partial Z_{i2}} = A(1-\gamma)\left(\gamma Z_{i1}^{-\delta} + (1-\gamma) Z_{i2}^{-\delta}\right)^{(-1/\delta)-1} Z_{i2}^{-\delta-1}.$$

Therefore

$$-\frac{S_1}{S_2} = \frac{\gamma}{(1-\gamma)}\left(\frac{Z_{i2}}{Z_{i1}}\right)^{\delta+1}, \tag{4}$$

where we can obtain estimates of $\frac{S_1}{S_2}$ directly from the estimated conditional quantiles, followed by a standard log linear regression for consistent estimation of γ and δ. In this method, unrestricted nonparametric distributions of the desired stitching variables are obtained, projected onto "equi-probable", quantiles, by means of estimated derivatives. The points in such quantile "sets" are then fitted to the aggregate functional as appropriate. This technique provides purely data driven values for the unknown parameters of the aggregator score function.

4.1.2 Stable Aggregation/Calibration

Though the two-step estimation is straightforward, the dimension of attributes is large in many economic application, e.g. return to schooling. When $d \geq 3$, the first step nonparametric estimation suffers from the curse of dimensionality. Thus we adopt a notion of stable aggregation: the weights and substitution degree between the component covariates in the aggregation function/indicator are the same in all cross section waves. This assumption is implicit in propensity score estimation, as well as in the synthetic control methods, and other data combination approaches. [6]

Denote the S_{i1} and S_{j2} as the aggregator functions two time points [7]:

$$
\begin{aligned}
S_{i1} &= A_1\left(\sum_{k=1}^{d} \gamma_j Z_{ik}^{-\delta}\right)^{-1/\delta}, i = 1, \cdots, n, \\
S_{j2} &= A_2\left(\sum_{k=1}^{d} \gamma_j Z_{jk}^{-\delta}\right)^{-1/\delta}, j = 1, \cdots, n.
\end{aligned} \tag{5}
$$

We choose γ's and δ to minimize the distance between the distributions of the two aggregator indexes, i.e., making S_{i1} and S_{j2}'s distributions as close as possible. We adopt the Hellinger distance as our optimization criterion. For details about the Hellinger distance, see Maasoumi and Racine (2002); Granger, Maasoumi, and Racine (2004). We normalize S_{i1} and S_{j2} to be in $[0, 1]$, i.e. $\sum_{i=1}^{n} S_{i1} = 1$ and $\sum_{j=1}^{n} S_{j2} = 1$. Let f_1 and f_2 be the density functions of S_{i1} and S_{j2} and the Hellinger distance is defined as

$$
S_\rho = \frac{1}{2}\int_0^1 \left(f_1^{1/2}(z) - f_2^{1/2}(z)\right)^2 dz. \tag{6}
$$

This method is akin to the method of "Hedonic Weights" for information aggregation, see Decancq and Lugo (2013). We choose the values of γ_j's and δ to minimize S_ρ to provide the the best match between entire samples.

4.2 Aggregation Algorithm Compared to Propensity Scores

Consider the aggregator function $S(z)$ and its CDF $F(s)$. Our stitching algorithm inverts based on the empirical counterpart of $F(s)$ to obtain the ranks of individuals (units), $F^{-1}(\hat{s}_i) = \hat{p}_i$ given $\hat{s}_i$ estimated aggregate for individual i. Units are then naturally ranked by $\hat{p}_i \in [0, 1]$. Units with the same (or similar) rank $\hat{p}_i$ in different periods are matched.

[6] This assumption can be relaxed to hold over subgroups, or sub periods.

[7] There are usually more than two periods, the result can be generalized accordingly.

In propensity score matching, an observed binary outcome $y \in \{0, 1\}$ is regressed on some CDF transform of (typically linear) index of variable z, such as Probit or Logit (or semiparametrically or nonparametrically), obtaining $F^*_{PS}(z_i'\hat{\beta}) = \hat{p}_i^*$[8]. Matching is then based on $\hat{p}_i^*$. There is other way to use propensity score for missing data, e.g., inverse probability weighting (see Graham, 2011; Graham, Pinto, and Egel, 2012).

There are two major differences. In the propensity score method, observations on a state $y \in \{0, 1\}$ is available, allowing a regression based predication of p_i^*. The inference theory for objects based on estimated propensity scores, such as average treatment effect, would be similar, as would robustness properties of doubly robust estimators in program evaluation. In particular, estimators based on estimated scores are non standard (non smooth) functionals, and standard bootstrap methods for them would be invalid (see Abadie and Imbens, 2008). Subsampling and sample splitting methods are required in these settings. We note in passing that see Wooldridge (2010). p_i^*'s admits the same "interpretation" as p_i's, as "assignment probabilities" to cohorts, for example. However, when an outcome (treatment choice) indicator is not available, our method still provides a "score" like the propensity score. But in this case, inferential functions are smooth functions of estimated scores, and standard asymptotic theory and bootstrap would be available. This is the case in the current paper and its applications. The FE estimator is a smooth function of observations. [9]

4.3 Comparison with Cohort Analysis

In the cohort approach a single (or two) variable provides for the "assignment rule", typically a time invariant variable, e.g. year of birth. We construct "cohorts" based on the constructed index $S_i(\cdot)$, which is not restricted to time invariant attributes. Under the common trend condition above, $S_i(\cdot)$ provides the basis for cohort designation. Further (since S_i lies between 0 and 1), we can conduct cluster analysis to determine optimal cohort size and the number of cohorts, as in Hirschberg, Maasoumi, and

[8] We use p^* to denote the score in propensity score matching.

[9] in ongoing work we consider empirical likelihood estimation of the unknown parameters in the aggregator function. These are known as Maximum Entropy estimators in information theory, with well developed inference methods.

Slottje (1991, 2001). The "donor" observations come from any desired cross section. This provides for multiple matches, if desired, and the number of cohorts is essentially a tuning parameter. Importantly, unlike the cohort approach, we do not employ cohort averages as observations to be attributed to cohort memebers. We are modelling actual unit observations, not cohort averages. Synthetic variable technique is closer to cohort method in this regard, than our method. Asymptotic analyses in the cohort literature allows the number of cohorts to increase to infinity, with the size of cohort going to infinity, or both. One may adapt the tuning parameter based on the cross sectional and time dimension considerations of the observed data.

5 Monte Carlo

We investigate the finite sample performance of our method based on the fixed effects estimator with artificially generated panel data. We generate full panel data draws from several data generating processes (DGPs). The sampled data is then treated as repeated cross sections, ignoring its longitudinal relations. We then apply our stitching method and employ the stitched pseudo panels.

We consider short panels with two periods and five periods. For each case, we use three different cross sectional dimensions, i.e., $n = 200, 400$, and 800. We work with balanced panels in the Monte Carlo studies. DGPs contains AR(1) sequences in the time dimension and the unobserved heterogeneity terms take means over time. DGP1, DGP2 and DGP3 contain three covariates: X_1 is constructed based on truncated standard normal distributed variables; X_2 is constructed based on exponentially distributed variables; and X_3 is constructed based on standard uniform distributed variables. We use the non-negative random variables because the nature of the optimal aggregator function, i.e., it cannot allow negative x due to its functional form[10]. For the five-period DGP2, we use a relatively strong persistent autocorrelation coefficient, $\rho = 0.9$. In DGP3, we introduce dependence among X_1, X_2 and X_3. We take the true

[10]In empirical studies, data transformation to positive observations can be performed before applying the stitching method.

parameters for the panel data model as $(\beta_1, \beta_2, \beta_3) = (-1, 1, -1)$ for all DGPs.

We then follow our proposed stitching procedure to take the genuine panels as if they are repeated cross sections. We use the calibration-type estimator for the coefficients in the optimal aggregator functions. Because we are dealing with balanced panels, we stitch individuals across periods based on their ranks presented in the estmated S's. For instance, the individual with lowest S in the first period is matched with the individual with lowest S in the second period, and so on. Fixed effects estimates obtain for both the original generated panels and the stitched pseudo panels. We generate 100 samles of psuedo-random numbers for illustration.

Our stitching estimator is compared with the FE estimator for the generated data using the ratio of the root mean squared errors (RRMSE$\equiv RMSE_{FE}/RMSE_{SE}$), computed around the true parameters. We report the results in Tables 1 to 3. In terms of RMSE, the performance of the stitching estimator is remarkably similar to the traditional FE estimator. The performance of our method is even better than the panel FE method in the strong persistent case. One potential reason is that our stitching procedures remove the persistence of individual's X's over the time.

DGP1:

$$
\begin{aligned}
U_{i1} &\sim \mathcal{N}(0,1)[0,\infty];\ U_{i2} \sim Exp(1);\ U_{i3} \sim U[0,1],\\
V_{i1} &\sim \mathcal{N}(0,1)[0,\infty];\ V_{i2} \sim Exp(1);\ V_{i3} \sim U[0,1],\\
X_{1,i1} &= U_{i1};\ X_{1,i2} = \rho X_{1,i1} + \sigma V_{i1},\\
X_{2,i1} &= U_{i2};\ X_{2,i2} = \rho X_{2,i1} + \sigma V_{i2},\\
X_{3,i1} &= U_{i3};\ X_{3,i2} = \rho X_{3,i1} + \sigma V_{i3},\\
\alpha_i &= \overline{X}_{1,i}/3 + \overline{X}_{2,i}/3 + \overline{X}_{3,i}/3 - (1+\rho+\sigma)/6 \cdot \Big(\sqrt{\frac{2}{\pi}} + 1 + 0.5\Big),\\
w &= 0.5 \times Std(\beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it}).\\
Y_{it} &= \alpha_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + U_{it},\ i = 1, \cdots, n; t = 1,2;\ \ U_{it} \sim \mathcal{N}(0, w^2).
\end{aligned}
$$

Table 1: Comparison of SE and FE Estimations (DGP1, $\rho = 0.8, \sigma = 0.2$)

Root Mean Square Errors									
n	RMSE of SE			RMSE of FE			RRMSE		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
200	0.332	0.236	0.474	0.302	0.166	0.664	0.909	0.701	1.402
400	0.251	0.182	0.385	0.216	0.156	0.431	0.861	0.856	1.120
800	0.167	0.116	0.257	0.153	0.095	0.389	0.918	0.820	1.513

DGP2:

$$U_{1i} \sim \mathcal{N}(0,1)[0,\infty];\ U_{2i} \sim Exp(1);\ U_{3i} \sim U[0,1]$$

$$V_{1,it} \sim \mathcal{N}(0,1)[0,\infty];\ V_{2,it} \sim Exp(1);\ V_{3,it} \sim U[0,1], t = 2,3,4,5$$

$$X_{1,i1} = U_{1i};\ X_{1,it} = \rho X_{1,it-1} + \sigma V_{1,it}, t = 2,3,4,5;$$

$$X_{2,i1} = U_{2i};\ X_{2,it} = \rho X_{2,it-1} + \sigma V_{2,it}, t = 2,3,4,5;$$

$$X_{3,i1} = U_{3i};\ X_{3,it} = \rho X_{3,it-1} + \sigma V_{3,it}, t = 2,3,4,5;$$

$$\alpha_i = \overline{X}_{1,i}/3 + \overline{X}_{2,i}/3 + \overline{X}_{3,i}/3 - (1 + \rho + \rho^2 + \rho^3 + \rho^4 + \rho^3\sigma + 2\rho^2\sigma + 3\rho\sigma + 4\sigma)/6 \cdot \left(\sqrt{\frac{2}{\pi}} + 1.5\right)$$

$$w = 0.5 \times Std(\beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it}).$$

$$Y_{it} = \alpha_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + U_{it},\ i = 1,\cdots,n; t = 1,2;\ \ U_{it} \sim \mathcal{N}(0, w^2),$$

Table 2: Comparison of SE and FE Estimations (DGP2, $\rho = 0.9, \sigma = 0.1$)

Root Mean Square Errors									
n	RMSE of SE			RMSE of FE			RRMSE		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
200	0.150	0.140	0.256	0.168	0.124	0.344	1.123	0.885	1.344
400	0.116	0.106	0.227	0.112	0.081	0.283	0.959	0.761	1.247
800	0.090	0.085	0.188	0.092	0.053	0.188	1.031	0.620	1.000

We then introduce dependence among X's in DGP3 for two-period case to investigate the finite sample performance of our proposed SE method. The optimal aggregator may draw different weights if there are dependence among X's. The results shown in Table 3 are similar to those from cases with independent X's.

DGP3:

$$
\begin{aligned}
&U_{i1} \sim \mathcal{N}(0,1)[0,\infty];\ U_{i2} \sim Exp(1);\ U_{i3} \sim U[0,1]; U_{i4} \sim \mathcal{N}(0,1)[0,\infty]\\
&V_{i1} \sim \mathcal{N}(0,1)[0,\infty];\ V_{i2} \sim Exp(1);\ V_{i3} \sim U[0,1]\\
X_{1,i1} &= 0.5U_{i1} + 0.5U_{i4};\ X_{1,i2} = \rho X_{1,i1} + \sigma V_{i1}\\
X_{2,i1} &= 0.5U_{i2} + 0.5U_{i4};\ X_{2,i2} = \rho X_{2,i1} + \sigma V_{i2}\\
X_{3,i1} &= 0.5U_{i3} + 0.5U_{i4};\ X_{3,i2} = \rho X_{3,i1} + \sigma V_{i3}\\
\alpha_i &= \overline{X}_{1,i}/3 + \overline{X}_{2,i}/3 + \overline{X}_{3,i}/3 - (1+\rho)/6 \cdot \left(2\sqrt{\frac{2}{\pi}} + 0.75\right) - \sigma/6 \cdot \left(\sqrt{\frac{2}{\pi}} + 1.5\right)\\
w &= 0.5 \times Std(\beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it}).\\
Y_{it} &= \alpha_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + U_{it},\ i = 1,\cdots,n; t = 1,2;\ \ U_{it} \sim \mathcal{N}(0,w^2).
\end{aligned}
$$

Table 3: Comparison of SE and FE Estimations (DGP1, $\rho = 0.8, \sigma = 0.2$)

Root Mean Square Errors									
n	RMSE of SE			RMSE of FE			RRMSE		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
200	0.320	0.275	0.382	0.213	0.140	0.378	0.667	0.509	0.990
400	0.224	0.233	0.339	0.165	0.101	0.249	0.737	0.434	0.735
800	0.207	0.206	0.300	0.106	0.076	0.171	0.512	0.367	0.571

6 Empirical Illustration

We illustrate our method in two genuine panel data applications. First is an examination of private return to R&D in the presence of spillovers in Eberhardt, Helmers, and Strauss (2013), using an unbalanced panel of up to 12 manufacturing subsectors in 10 OECD countries, over a maximum of a 26-year period. The second is concerned with female labor force participation from Fernández-Val (2009) using a balanced panel of

1461 females from the PSID dataset. We treat the data in both cases ignoring their panel structure, as if they are repeated cross sections. We apply our stitching approach to obtain balanced "pseudo" panels, and use the same estimation methods employed in the original Eberhardt, Helmers, and Strauss (2013) and Fernández-Val (2009) studies. We compare estimation results and find generally qualitatively similar inferences.

6.1 Private Return to R&D in the Presence of Spillovers

Examination of private returns to R&D in the presence of spillovers was conducted based on real panels by Eberhardt, Helmers, and Strauss (2013) and Millo (2019). Stitching was done on the basis of three variables, logrithm labor ($\ln L_{it}$), logrithm capital ($\ln K_{it}$) and logrithm R&D ($\ln R_{it}$) over the 26 years. The original data is an unbalanced panel. We elongate the data by stitching units with closest ranked units across time for a balanced stitched pseudo panel of 118 units over 26 periods (using the calibration technique for estimating aggregators). We explore the four different specifications adopted in Eberhardt, Helmers, and Strauss (2013) for estimation: static and dynamic homogenous models, and static and dynamic heterogenous models. The original estimates are reported in the left panel and our estimates in the right panel in Tables 4 and 5.

The estimates of private returns to R&D in the original paper and our proposed method are qualitatively similar in homogenous models. In heterogenous models, private returns to R&D are all positive based on our stitching method, whereas some effects are negative in the original paper.

Based on our stitched pseudo panels, accounting for cross sectional dependence does not make a large difference in estimates for private returns to labor, capital and R&D. This may be due to that stitching removes the spillovers of knowledge, which leads to more independent "pseudo" panels waves. This may be different if our methods are used to impute "missing observations" or missing responses in otherwise real panels.

Homogenous Models (Static)										
	Eberhardt, Helmers, and Strauss (2013)					Stitching Estimation				
	POLS	2FE	FD	CCEP	CCEPt	POLS	2FE	FD	CCEP	CCEPt
$\ln L_{it}$	0.464	0.608	0.635	0.562	0.582	0.464	0.562	0.468	0.525	0.525
t-statistics	40.946	18.944	18.085	20.714	21.002	44.169	14.112	5.509	5.762	5.737
$\ln K_{it}$	0.465	0.487	0.279	0.289	0.203	0.476	0.484	0.449	0.468	0.468
t-statistics	37.802	10.908	3.431	7.946	4.972	40.983	40.279	28.095	38.780	38.616
$\ln R_{it}$	0.096	0.063	0.045	0.084	0.064	0.082	0.078	0.085	0.078	0.078
t-statistics	22.923	4.544	1.698	4.925	3.662	19.932	18.138	14.664	19.961	19.876
Homogenous Models (Dynamic)										
	Eberhardt, Helmers, and Strauss (2013)					Stitching Estimation				
	POLS	2FE	BB	CCEP	CCEPt	POLS	2FE	BB	CCEP	CCEPt
$\ln L_{it}$	0.338	0.654	-0.792	0.369	0.364	0.440	0.535	0.444	0.321	0.321
t-statistics	2.495	19.761	-0.927	5.256	4.996	22.885	9.482	3.670	2.034	2.025
$\ln K_{it}$	0.173	0.078	1.409	0.367	0.287	0.492	0.461	0.489	0.496	0.496
t-statistics	0.869	1.447	2.041	3.746	2.644	22.879	40.914	5.357	22.419	22.324
$\ln R_{it}$	0.462	0.019	0.222	0.066	0.064	0.091	0.083	0.054	0.081	0.081
t-statistics	2.788	0.815	1.594	1.668	1.600	12.427	22.388	1.494	11.432	11.383

Table 4: Static and Dynamic homogeneous models (Table 5 and 6 in EHS).

Heterogenous Models (Static)								
	Eberhardt, Helmers, and Strauss (2013)				Stitching Estimation			
	MG	CDMG	CMG	CMGt	MG	CDMG	CMG	CMGt
$\ln L_{it}$	0.568	0.557	0.599	0.698	0.811	0.701	0.715	0.694
t-statistics	6.569	7.628	9.000	8.236	9.471	7.230	4.787	4.013
$\ln K_{it}$	0.117	0.445	0.244	0.149	0.431	0.434	0.418	0.410
t-statistics	0.955	5.008	1.702	1.004	13.448	10.113	12.524	11.249
$\ln R_{it}$	-0.058	0.089	0.035	-0.050	0.070	0.066	0.075	0.075
t-statistics	-0.728	2.123	0.445	-0.601	7.458	4.958	7.034	6.938
Heterogeous Models (Dynamic)								
	Eberhardt, Helmers, and Strauss (2013)				Stitching Estimation			
	MG	CDMG	CMG	CMGt	MG	CDMG	CMG	CMGt
$\ln L_{it}$	0.703	0.567	0.642	0.678	0.865	0.741	0.856	0.987
t-statistics	6.152	10.011	9.386	9.432	8.314	7.332	3.720	4.240
$\ln K_{it}$	0.277	0.245	0.276	0.172	0.408	0.403	0.401	0.417
t-statistics	1.867	3.373	1.709	1.088	14.484	14.537	10.524	9.221
$\ln R_{it}$	-0.107	0.139	-0.084	-0.088	0.080	0.088	0.072	0.061
t-statistics	-0.953	3.947	-0.945	-0.964	8.126	9.847	4.985	4.145

Table 5: Heterogeneous models (Table 7 and 8in EHS).

6.2 Female Labor Force Participation

In the study of female labor force participation, Fernández-Val (2009) used PSID Waves 13-22, with 1461 females over 10 years. Fernández-Val proposed a bias correction method to handle the incidental parameters problem in the nonlinear panel data model. He examined the relationship between fertility and female labor force participation using a nonlinear panel data model with unobserved heterogeneity to deal with multiple unobserved factors as determinants of joint fertility and female labor force participation decisions. We take the micropanel as if it is repeated cross sections and use our stitching method to generate a stitched pseudo panel based on the calibration estimation of the aggreagator function. Stitching variables are logrithm of husband's income and age. Applying the same estimation strategies used in Fernández-Val (2009), we obtain similar results on the impact of fertility on female labor force participation. We report the result for the Logit and Probit methods in Tables 6 to 8. The results of linear probability model are similar. Our results are also consistent with the finding of Fernández-Val (2009) that uncorrected estimates of index coefficients are larger (in absolute value) than their bias-corrected counterparts. The signs and significance of all estimates in both the static and the dynamic models are the same as those found in Fernández-Val (2009).

7 Conclusion

In this paper, we examine the performance of a pseudo panel construction based on an optimization of whole sample imputation technique. A rank preservation condition on unobserved heterogeneity helps to transform time series of cross sections to pseudo panels which retain the attractive time invariant heterogeneity feature in a genuine panel. The stitched pseudo panels so constructed allow traditional fixed effects inferences. We do not use averages of cohorts for imputation of missing or unobserved objects. Our approach has many other applications in similar situations, including pure cross section-treatment effect applications, as in Maasoumi and Eren

	Fernández-Val (2009)				Stitching Estimation			
Estimator	FE	JK	BC3	BC3p	FE	JK	BC3	BC3p
A- Index Coefficients								
Kids 0-2	-0.714	-0.618	-0.631	-0.666	-0.525	-0.447	-0.467	-0.460
	(0.056)	(0.055)	(0.055)	(0.060)	(0.031)	(0.031)	(0.031)	(0.032)
Kids 3-5	-0.411	-0.363	-0.364	-0.382	-0.309	-0.261	-0.275	-0.274
	(0.051)	(0.051)	(0.051)	(0.055)	(0.028)	(0.028)	(0.028)	(0.027)
Kids 6-17	-0.130	-0.102	-0.115	-0.128	-0.085	-0.072	-0.075	-0.073
	(0.041)	(0.041)	(0.041)	(0.046)	(0.014)	(0.014)	(0.014)	(0.014)
Log(Husband income)	-0.242	-0.210	-0.214	-0.210	-0.179	-0.154	-0.160	-0.160
	(0.054)	(0.053)	(0.053)	(0.056)	(0.020)	(0.020)	(0.020)	(0.022)
B- Marginal Effects (%)								
Kids 0-2	-9.279	-9.474	-9.135	-9.636	-14.941	-14.869	-14.748	-14.529
	(0.697)	(0.704)	(0.702)	(0.764)	(0.869)	(0.880)	(0.877)	(0.892)
Kids(3-5	-5.344	-5.518	-5.264	-5.529	-8.788	-8.718	-8.682	-8.658
	(0.656)	(0.661)	(0.660)	(0.708)	(0.788)	(0.797)	(0.794)	(0.783)
Kids 6-17	-1.687	-1.602	-1.665	-1.851	-2.413	-2.389	-2.385	-2.304
	(0.532)	(0.537)	(0.536)	(0.598)	(0.399)	(0.402)	(0.402)	(0.401)
Log(Husband income)	-3.140	-3.195	-3.098	-3.036	-5.099	-5.113	-5.048	-5.049
	(0.695)	(0.699)	(0.698)	(0.735)	(0.575)	(0.578	(0.577)	(0.635)

Table 6: Static model [Table 10 (Probit Part)] in Fernández-Val (2009).

	Fernández-Val (2009)					Stitching Estimation				
Estimator	FE	JK	BC3	C	BC3p	FE	JK	BC3	C	BC3p
A- Index Coefficients										
Kids 0-2	-0.683	-0.591	-0.599	-0.599	-0.631	-0.486	-0.417	-0.427	-0.427	-0.420
	(0.054)	(0.053)	(0.053)	(0.053)	(0.057)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Kids 3-5	-0.393	-0.346	-0.345	-0.345	-0.362	-0.286	-0.244	-0.252	-0.252	-0.251
	(0.049)	(0.049)	(0.049)	(0.049)	(0.051)	(0.026)	(0.026)	(0.026)	(0.026)	(0.025)
Kids 6-17	-0.129	-0.106	-0.114	-0.114	-0.126	-0.081	-0.071	-0.072	-0.072	-0.069
	(0.039)	(0.039)	(0.039)	(0.039)	(0.043)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Log(Husband income)	-0.229	-0.199	-0.202	-0.202	-0.198	-0.174	-0.155	-0.154	-0.154	-0.154
	(0.052)	(0.051)	(0.051)	(0.051)	(0.054)	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)
B- Marginal Effects (%)										
Kids 0-2	-9.414	-9.459	-9.260	-9.257	-9.768	-14.837	-14.74	-14.583	-14.578	-14.368
	(0.715)	(0.717)	(0.715)	(0.715)	(0.760)	(0.864)	(0.873)	(0.872)	(0.872)	(0.879)
Kids 3-5	-5.414	-5.493	-5.341	-5.332	-5.598	-8.733	-8.655	-8.617	-8.619	-8.591
	(0.667)	(0.670)	(0.669)	(0.669)	(0.706)	(0.780)	(0.787)	(0.786)	(0.786)	(0.776)
Kids 6-17	-1.783	-1.728	-1.766	-1.762	-1.946	-2.475	-2.488	-2.446	-2.450	-2.363
	(0.543)	(0.546)	(0.546)	(0.546)	(0.591)	(0.400)	(0.402)	(0.402)	(0.402)	(0.402)
Log(Husband income)	-3.160	-3.176	-3.121	-3.120	-3.063	-5.304	-5.430	-5.262	-5.269	-5.269
	(0.710)	(0.710)	(0.709)	0.709)	(0.743)	(0.589)	(0.590)	(0.590)	(0.590)	(0.633)

Table 7: Static model [Table 10 (Logit Part)] in Fernández-Val (2009).

(2006), or time series applications as in Ginindza and Maasoumi (2013) who analysed Difference-in-Difference effects of inflation targetting for a sample of countries. The

	Fernández-Val (2009)				Surrogate Estimation			
Estimator	Probit		Logit		Probit		Logit	
	FE	BC3	FE	BC3	FE	BC3	FE	BC3
A- Index Coefficients								
Participation$_{t-1}$	0.756	1.031	0.693	0.944	2.489	2.041	2.430	1.892
	(0.042)	(0.043)	(0.039)	(0.041)	(0.041)	(0.030)	(0.046)	(0.030)
Kids 0-2	-0.554	-0.436	-0.531	-0.418	-0.322	-0.251	-0.337	-0.266
	(0.061)	(0.062)	(0.058)	(0.059)	(0.045)	(0.037)	(0.045)	(0.036)
Kids 3-5	-0.279	-0.193	-0.264	-0.182	-0.104	-0.085	-0.096	-0.081
	(0.055)	(0.058)	(0.052)	(0.054)	(0.038)	(0.031)	(0.039)	(0.031)
Kids 6-17	-0.075	-0.050	-0.074	-0.050	-0.004	0.003	-0.004	0.004
	(0.045)	(0.045)	(0.042)	(0.043)	(0.018)	(0.015)	(0.019)	(0.015)
Log(Husband income)	-0.246	-0.209	-0.234	-0.199	-0.130	-0.104	-0.136	-0.110
	(0.056)	(0.057)	(0.054)	(0.054)	(0.025)	(0.021)	(0.026)	(0.021)
B- Marginal Effects (%)								
Participation$_{t-1}$	10.724	17.097	10.492	17.202	65.296	64.381	65.437	64.025
	(0.640)	(0.667)	(0.633)	(0.660)	(0.743)	(0.750)	(0.740)	(0.749)
Kids 0-2	-6.796	-5.962	-6.851	6.009	-5.142	-5.007	-5.317	-5.330
	(0.739)	(0.725)	(0.736)	(0.720)	(0.718)	(0.699)	(0.699)	(0.695)
Kids 3-5	-3.426	-2.637	-3.403	-2.613	-1.669	-1.685	-1.518	-1.618
	(0.677)	(0.670)	(0.673)	(0.665)	(0.613)	(0.597)	(0.619)	(0.609)
Kids 6-17	-0.919	-0.687	-0.958	-0.724	-0.059	0.051	-0.058	0.077
	(0.546)	(0.537)	(0.537)	(0.528)	(0.295)	(0.288)	(0.301)	(0.296)
Log(Husband income)	-3.020	2.853	-3.020	-2.859	-2.071	-2.074	-2.141	-2.215
	(0.689)	(0.672)	(0.691)	(0.668)	(0.406)	(0.401)	(0.414)	(0.413)

Table 8: Dynamic model [Table 11, Probit and Logit] in Fernández-Val (2009).

performance of our proposed approach appears to be quite satisfactory. We conjecture that this is due to optimum information processing by our aggregation method, and the emphasis on stitching entire samples.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association 105*(490), 493–505.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review 93*(1), 113–132.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. W. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557.

Abadie, A. and G. W. Imbens (2016). Matching on the estimated propensity score. *Econometrica 84*(2), 781–807.

Andersen, E. B. (1973). *Conditional inference and models for measuring*, Volume 5. Mentalhygiejnisk forlag.

Arellano, M. (2003). *Panel data econometrics*. Oxford university press.

Arellano, M. and S. Bonhomme (2011). Nonlinear panel data analysis. *Annual Review of Economics 3*(1), 395–424.

Arellano, M. and S. Bonhomme (2017). Nonlinear panel data methods for dynamic heterogeneous agent models. *Annual Review of Economics 9*(1), 471–496.

Arellano, M. and R. Carrasco (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics 115*(1), 125–157.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs 43*, 381.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of Econometrics 5*, 3229–3296.

Attanasio, O. P. (1998). Cohort analysis of saving behavior by us households. *Journal of Human Resources*, 575–609.

Attanasio, O. P., J. Banks, C. Meghir, and G. Weber (1999). Humps and bumps in lifetime consumption. *Journal of Business & Economic Statistics 17*(1), 22–35.

Attanasio, O. P. and G. Weber (1993). Consumption growth, the interest rate and aggregation. *The Review of Economic Studies 60*(3), 631–649.

Attanasio, O. P. and G. Weber (1995). Is consumption growth consistent with intertemporal optimization? evidence from the consumer expenditure survey. *Journal of Political Economy 103*(6), 1121–1157.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J. and S. Ng (2019). Fbi: Factor based imputation of missing values. *Working paper*.

Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

Bester, C. A. and C. Hansen (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics 27*(2), 131–148.

Blundell, R., M. Browning, and C. Meghir (1994). Consumer demand and the life-cycle allocation of household expenditures. *The Review of Economic Studies 61*(1), 57–80.

Blundell, R., A. Duncan, and C. Meghir (1998). Estimating labor supply responses using tax reforms. *Econometrica*, 827–861.

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

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics 3*, 1801–1863.

Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics 140*(2), 503–528.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies 47*(1), 225–238.

Chamberlain, G. (1984). Panel data. *Handbook of Econometrics 2*, 1247–1318.

Chen, M., I. Fernández-Val, and M. Weidner (2019). Nonlinear factor models for network and panel data. *arXiv preprint arXiv:1412.5647*.

Chen, X., H. Hong, and A. Tarozzi (2008). Semiparametric efficiency in gmm models with auxiliary data. *The Annals of Statistics 36*(2), 808–843.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological) 20*(2), 215–232.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics 30*(1-2), 109–126.

Decancq, K. and M. A. Lugo (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews 32*(1), 7–34.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Eberhardt, M., C. Helmers, and H. Strauss (2013). Do spillovers matter when estimating private returns to r&d? *The Review of Economics and Statistics 95*(2), 436–448.

Fan, Y., R. Sherman, and M. Shum (2014). Identifying treatment effects under data combination. *Econometrica 82*(2), 811–822.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics 150*(1), 71–85.

Fernández-Val, I. and F. Vella (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics 163*(2), 144–162.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics 192*(1), 291–312.

Fernández-Villaverde, J. and D. Krueger (2007). Consumption over the life cycle: Facts from consumer expenditure survey data. *The Review of Economics and Statistics 89*(3), 552–565.

Fernandez-Villaverde, J. and D. Krueger (2011). Consumption and saving over the life cycle: How important are consumer durables? *Macroeconomic dynamics 15*(5), 725–770.

Gayer, G., O. Lieberman, and O. Yaffe (2019). Similarity-based model for ordered categorical data. *Econometric Reviews 38*(3), 263–278.

Gilboa, I., O. Lieberman, and D. Schmeidler (2006). Empirical similarity. *The Review of Economics and Statistics 88*(3), 433–444.

Gilboa, I., O. Lieberman, and D. Schmeidler (2011). A similarity-based approach to prediction. *Journal of Econometrics 162*(1), 124–131.

Gilboa, I. and D. Schmeidler (2003). Inductive inference: An axiomatic approach. *Econometrica 71*(1), 1–26.

Ginindza, M. and E. Maasoumi (2013). Evaluating inflation targeting based on thedistribution of inflation and inflation volatility. *North American Journal of Economics and FinanceEvaluating 26*, 497–518.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica 79*(2), 437–452.

Graham, B. S., C. C. d. X. Pinto, and D. Egel (2012). Inverse probability tilting for moment condition models with missing data. *The Review of Economic Studies 79*(3), 1053–1079.

Graham, B. S., C. C. d. X. Pinto, and D. Egel (2016). Efficient estimation of data combination models by the method of auxiliary-to-study tilting (ast). *Journal of Business & Economic Statistics 34*(2), 288–301.

Granger, C., E. Maasoumi, and J. Racine (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis 25*(5), 649–669.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal 7*(1), 98–119.

Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory 27*(6), 1152–1191.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458.

Hellerstein, J. K. and G. W. Imbens (1999). Imposing moment restrictions from auxiliary data by weighting. *Review of Economics and Statistics 81*(1), 1–14.

Hirano, K., G. W. Imbens, G. Ridder, and D. B. Rubin (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica 69*(6), 1645–1659.

Hirschberg, J. G., E. Maasoumi, and D. J. Slottje (1991). Cluster analysis for measuring welfare and quality of life across countries. *Journal of Econometrics 50*(1-2), 131–150.

Hirschberg, J. G., E. Maasoumi, and D. J. Slottje (2001). Clusters of attributes and well-being in the usa. *Journal of Applied Econometrics 16*(3), 445–460.

Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica 68*(4), 839–874.

Honoré, B. E. and A. Lewbel (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica 70*(5), 2053–2063.

Hsiao, C. (2014). *Analysis of panel data*. Cambridge university press.

Inoue, A. and G. Solon (2010). Two-sample instrumental variables estimators. *The Review of Economics and Statistics 92*(3), 557–561.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica 54*(4), 991–997.

Maasoumi, E. and O. Eren (2006). The information basis of matching with propensity scores. *Working paper*.

Maasoumi, E. and J. Racine (2002). Entropy and predictability of stock market returns. *Journal of Econometrics 107*(1), 291–312.

Maasoumi, E. and J. Racine (2016). A solution to aggregation and an application to multidimensional "well-being" frontiers. *Journal of Econometrics 191*(2), 374–383.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy 89*(6), 1059–1085.

Millo, G. (2019). Private returns to r&d in the presence of spillovers, revisited. *Journal of Applied Econometrics 34*(1), 155–159.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics 59*(1-2), 99–123.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*(4), 1543–1579.

Pagel, M. (2017). Expectations-based reference-dependent life-cycle consumption. *The Review of Economic Studies 84*(2), 885–934.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests.

Ridder, G. and R. Moffitt (2007). The econometrics of data combination. *Handbook of econometrics 6*, 5469–5547.

Riesbeck, C. K. and R. C. Schank (1989). *Inside case-based reasoning*. Lawrence Erlbaum Associates.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively*. Lawrence Erlbaum Associates.

Su, L. and S. Jin (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics 169*(1), 34–47.

Su, L., K. Miao, and S. Jin (2019). On factor models with random missing: Em estimation, inference, and cross validation. *Working paper*.

Su, L. and A. Ullah (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of empirical economics and finance*, 455–497.

Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. *The Econometrics of Panel Data*, 369–383.

Verbeek, M. and F. Vella (2005). Estimating dynamic models from repeated cross-sections. *Journal of econometrics 127*(1), 83–102.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.