Elongated Pseudo-Panels from Time Series of Cross Sections*

Zhongjian Lin† Esfandiar Maasoumi‡

Abstract

We propose to impute missing components in panels or entire waves in time series of cross sections. An "optimal matching" procedure is proposed to infer would-be multiple observations, accounting for unobserved heterogeneity. Matching is done by "aggregation" and extrapolation over any desired group of observed covariates. Our aggregation is information efficient and is applicable both with observed outcomes (such as in propensity score matching), or without. This includes a substantive generalization of "cohort" methods and restores the continuity of unobserved heterogeneity across time. This feature mitigates the major limitation of pseudo panel data that lack individual histories. Our synthetic panels are examined by simulation and in two empirical studies, on the private return to R&D in the presence of spillovers using macropanel data and female labor force participation using micropanel data (PSID) where we know the entire panel.

Keywords: Inductive Inference, Unobserved Heterogeneity, Pseudo Panels, Information Aggregation.

JEL Classifications:

1 Introduction

Panel data allow for control of unobserved heterogeneity and facilitate identification and estimation of structural parameters in presence of fixed effects and unobserved heterogeneity. See Chamberlain (1984), Arellano and Honoré (2001), Arellano (2003), Baltagi (2008), Wooldridge (2010), Arellano and Bonhomme (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 for short time periods. Short time micropanels are inadequate for longer life-cycle dynamic analyses. Furthermore, attrition can dramatically shrink the cross sectional dimension. Less expensive long

*This version: June 14, 2019
†Department of Economics, Emory University. Email: zhongjian.lin@emory.edu.
‡Department of Economics, Emory University. Email: esfandiar.maasoumi@emory.edu.

time series of cross sections, such as the CPS, can mitigate the short period and the attrition issues of panel data, but lack the desired repeated observations benefits of real panels.

In this paper we explore a new approach to creating panels from pseudo panels. A special situation that is similar in motivation, and subsumed by our general non-determinsitic methods, is the cohort analysis. 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 one characteristic (such as age) in these studies. This presumes a great deal of homogeneity in all other characteristics and dimensions, observed and otherwise. We should add, the cohort approach in studies of life-cycle consumption, saving and labor supply, is profligate, 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.

The cohort technique is an aggregation or grouping method which employs cohort averages. It is well known that averaging in this manner fails to preserve important heterogeneity features within groups and/or nonlinear effects and relations. In practice, the cohort assignment is restricted to a single time-invariant variable, and the cohort level covariates are required to have sufficient variation. Furthermore, in the instrumental variables analyses of the cohort context , see Moffitt (1993), 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 (2008). Seawright (2009) considered cohort matching and estimation of missing responses based on more than one covariate.

We propose a statistical maching technique that infers "similarity" of cross section units over different periods based on ideal aggregation of many observed characteristics and covariates in many dimensions. This is consistent with the modelling context in which observed hetergeneity is acknowledged in all other aspects of the model and underlying theory. Our approach requires a familiar assumption on rank similarity between the constructed "aggregate score" and the unobserved heterogeneity.

Extrapolation of similarity based on observables is fundamental to learning. "From causes which appear similar we expect similar effects. This is the sum of all our experimental conclusions.” This mode of thinking is common to methods of analogy, reasoning by similarities, or case-based reasoning in psychology, artificial intelligence, treatment effects and program evaluation [1].

In this paper, we propose ideal aggregates/indices of observed covariate for individual units, and match them with similar indices in different time periods. Missing waves of a would be panel data set are thus constructed. The missing waves are counterfactual cross sections and distributions, not of outcomes, but of aggregated score functions. We do not match cross section units by their outcomes. Our criteria for matching measures closeness of the **entire samples**, of scores, emphasizing imputation of **missing samples**. The performance of the approach is examined both in simulation experiments and by application to fixed effect (FE) models for known panels such as the PSID. In this paper we will focus on one-to-one matching, but multiple matching/imputation 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 demonstrates the optimal information aggregation of multivariate characteristics. Estimation of the parameters in the optimal aggregator is also discussed in Section 4. Section 5 provides simulation evidence, and Section 6 presents two empirical applications to private return to R&D in the presence of spillovers and 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, $\{Y_{it}, X_{it}\}_{i=1,\cdots,n;t=1,\cdots,T}$. The synthetic panel would maintain the desired feature that the ith individual possesses the same (or similar) unobserved heterogeneity. We illustrate in three examples

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

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

[1] 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.

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:

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

where ε follows standard Logistic distribution. A consistent estimator for structural parameter, β obtains from conditional MLE, see §23.4.3 of Cameron and Trivedi (2005) and §7.3.1.2 of Hsiao (2014) for details. 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).

For random coefficient panel data model, the applicability of our synthetic panel construction requires further investigation of the similarity of the random coefficients and is deferred to future research.

3 Analysis of Unobserved Heterogeneity

Accounting for unobserved individual heterogeneity often leads to significant and sunbstantively different practical inferences. Fixed effects approaches in panel data models are attractive due to their flexible treatment of unobserved heterogeneity; see Blundell and Stoker (2005, 2007), Blundell, MaCurdy, and Meghir (2007).

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 $S(\cdot)$[2].

***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 "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 extrapolate the rank of ability. It is clear that one does not have to employ all the observed characteristics in aggregation[3]. There are applications in which the rank similarity condition in Assumption 1 is not compelling.

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

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

Consider a series of independent cross sections, $\{Y_{i1}, X_{i1}\}_{i=1,\cdots,n}$ and $\{Y_{i'2}, X_{i'2}\}_{i'=1,\cdots,n}$[4]. 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 equivalent where $F_\alpha(\cdot)$ and $F'_\alpha(\cdot)$ are the distribution functions of α_i and $\alpha_{i'}$ respectively. Waves of cross sections usually

[2]We defer the discussion of "appropriate" information aggregation to Section 4. Single variate cohort assignment IS an aggregation method!

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

[4]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.

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 pseudo panel feasible. We summarize the above as a proposition:

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

***Remark* 2.** Proposition 1 may be generalized to multiple-to-one matching, 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 (2017).

Another similar situation is in propensity score matching in treatment effects studies, e.g. Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1998), Abadie and Imbens (2006, 2016). The aggregation function $S(\cdot)$ is similar to the propensity score function, but with a substative and substantial difference (see below). Further, the components of $S(\cdot)$ need not be time-invariant in our approach. The common trend condition is sufficient to preserve the rankings.

4 Information Aggregation of Observed Characteristics

In this section, we describe first an optimization approach to construction of ideal aggregates/score functions of the desired characteristics. The optimal aggregation is for each period and we suppress the subscript t when it is not confusing. In contrast to estimation of propensity scores, either in parametric or nonparametric ways, we do not have/employ an outcome (left hand side treatment) variable. A probit or logit type estimation is not feasible. In our approach matching is based only on observed characteristics.

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

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

$$S(Z_i) = \underset{S}{argmin}\ D_\delta(S,Z;\gamma) = \sum_{j=1}^{d}\gamma_j\left[\sum_{i=1}^{n} S_i\left[\left(S_i/Z_{ij}\right)^{\delta} - 1\right]/\delta(\delta+1)\right]\ s.t. \sum_{i=1}^{n} S_i = 1. \quad (4.1)$$

We have the optimal aggregator as

$$\begin{aligned} S_i &\propto \left(\sum_{j=1}^{d}\gamma_j Z_{ij}^{-\delta}\right)^{-1/\delta}, \quad \delta \neq 0,-1; \\ S_i &\propto \Pi_{j=1}^{d} Z_{ij}^{\gamma_j}, \quad \delta = 0; \\ S_i &\propto \sum_{j=1}^{d}\gamma_j Z_{ij}, \quad \delta = -1. \end{aligned} \quad (4.2)$$

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. We emphasize, this makes entire distributions (samples) closest to each other.

4.1 Estimation of Parameters in the Optimal Aggregator

As with propensity scores, we require estimates for the unknown parameters in these functions. 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 (4.2) with $\delta \neq 0,-1$.

4.1.1 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

$$S(Z_i) = A\left(\gamma Z_{i1}^{-\delta} + (1-\gamma)Z_{i2}^{-\delta}\right)^{-1/\delta}. \quad (4.3)$$

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

$$-\frac{S_1}{S_2} = \frac{\gamma}{(1-\gamma)} \left(\frac{Z_{i2}}{Z_{i1}} \right)^{\delta+1}, \tag{4.4}$$

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 matching 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 deriven values for the unknown parameters of the aggregator score function.

4.1.2 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 so-called curse of dimensionality. Thus we adopt calibration like method to estimate the parameters. The technique is similar to that in Maasoumi and Xu (2015). Maasoumi and Xu (2015) compare the distributions of the optimal aggregator (for happiness) and the reported happiness from the data and obtain those estimates which minimizes the distance of the two distributions, e.g. the Hellinger distance.

In our pseudo-panel framework, we do not utilize observed outcomes corresponding to the aggregator index. Fortunately, the representativeness of the pseudo panel facilitates a calibration avenue. Denote the S_{i1} and S_{j2} as the two aggregator functions for period 1 and 2[5]:

$$
\begin{aligned}
S_{i1} &= A_1\left(\sum_{k=1}^{d}\gamma_j Z_{ik}^{-\delta}\right)^{-1/\delta}, i = 1,\cdots,n, \\
S_{j2} &= A_2\left(\sum_{k=1}^{d}\gamma_j Z_{jk}^{-\delta}\right)^{-1/\delta}, j = 1,\cdots,n.
\end{aligned} \tag{4.5}
$$

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 Matusita-Bhattacharya-Hellinger distance as our optimization criterion. For details about the Hellinger distance, see Maasoumi and Racine (2002), Granger et al. (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

$$
S_\rho = \frac{1}{2}\int_0^1 \left(f_1^{1/2}(z) - f_2^{1/2}(z)\right)^2 dz. \tag{4.6}
$$

We search the whole parameter space to find the set of parameters, seeking the smallest distance **between entire samples**. The calibration method is akin to 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. Cross section units are not matched individually.

This sample matching approach is clearly more suited when statistical inference on models is the inferential objective. "Individual" matching may be better suited when treatment effect at individual level is sought.

4.2 Matching Algorithm Compared to Propensity Scores Methods

Consider the aggregator function $S(z)$ and its CDF $F(s)$, our matching 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 $p_i \in [0,1]$. Units with the same rank p_i in different periods are matched.

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

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

(or semiparametrically or nonparametrically), obtaining $F^*_{PS}(z_i'\hat{\beta}) = \hat{P}^*_i$. Matching is then based on $\hat{P}^*_i$.

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 $\hat{p}^*_i$. We do not have the state variable, and derivation of $\hat{s}_i$ is different– it is an essentially imputation method, as in our calibration or two-stage methods.

The second difference is that our $\hat{p}_i$'s are "ordered" and $\sum_i \hat{P}_i = 1$.

Exploration of inverse probability weighting and conditioning based on PS is beyond the scope of this paper, but such methods are clearly available based on $\hat{P}_i$, especially doubly robust inference techniques, which are tolerant of misspecified models that obtain these scores. See Wooldridge (2010). P^*_i's admits the same "interpretation" as P_i's, as "assignment probabilities" to cohorts, for example.

4.3 Cohort Analysis Revisited

The cohort approach may be interpreted as a very limited special case of matching, both in terms of its single variable "assignment rule", and in terms of follow up inference procedures. Cohort designation is typically based on 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, or any one variable. 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 Slottje (1991, 2001). This provides for multiple matches, if desired, and the number of cohorts is essentially a tuning parameter. 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.

Similar to single variable cohort methods, the aggregate scores S_i are less heterogeneous, and less variable, generally speaking. The assumption that $\bar{\alpha}_{ct} = \alpha_c$ is more plausible. This is important for all cohort analyses, especially when the time dimension is small, see Verbeek (2008) for a critical discussion of this assumption.

5 Monte Carlo

We investigate the performance of our methods 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 matching methods and employ the constructed pseudo panels (calibrated) to estimate parameters. These estimators are then compared with corresponding estimators from the genuine panel data.

We consider the short panels with two periods and five periods. For the two periods case, we generate genuine data for two periods using DGP with known parameters. The S scores for the first and second period are then computed. The matched pseudo panels are constructed based on the ranks of individuals from these S's. The individual with lowest S in the first period is matched with the individual with lowest S in the second period, and so on. Implicit empirical probability transforms provide the matches. The estimators are compared in terms of mean squared errors (MSE)fit for short panels.

Our pseudo (imputed) panel estimator is compared with the FE estimator for the above generated data using ratio of root mean squared errors (RRMSE$\equiv RMSE_{FE}/RMSE_{IE}$), computed around the true parameters $(\beta_1, \beta_2, \beta_3) = (-1, 1, -1)$. In the first DGP, we consider three independent X's. In the second DGP, we consider three dependent X's which may have different optimal weights in the CES optimal aggregator functions compared to the independent case. In terms of RMSE, the performance of the pseudo panel FE estimator is remarkably similar to the traditional FE estimator. In the third DGP, we consider five periods and there is persistence in observed characteristics. The performance of our method is even better that the panel FE method in the non-stationary case. One potential reason is that our matching removes the persistence of individual's X's over the time.

DGP1:

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

Table 1: Comparison of IE and FE Estimations (DGP1, $\rho = 0.7, \sigma = 0.3$)

Root Mean Square Errors									
n	RMSE of IE			RMSE of FE			RRMSE		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
200	0.411	0.311	0.602	0.415	0.238	0.831	1.010	0.768	1.380
400	0.327	0.241	0.456	0.286	0.180	0.621	0.874	0.747	1.362
800	0.199	0.164	0.301	0.195	0.141	0.399	0.978	0.856	1.325

DGP2:

$$
\begin{aligned}
& U_{1i} \sim \mathcal{N}(0,1)[0,\infty];\ U_{2i} \sim Exp(1);\ U_{3i} \sim U[0,1] \\
& V_{1,it} \sim \mathcal{N}(0,1)[0,\infty];\ V_{2,it} \sim Exp(1);\ V_{3,it} \sim U[0,1], t = 2,3,4,5 \\
& X_{1,i1} = U_{1i};\ X_{1,it} = \rho X_{1,it-1} + \sigma V_{1,it}, t = 2,3,4,5; \\
& X_{2,i1} = U_{2i};\ X_{2,it} = \rho X_{2,it-1} + \sigma V_{2,it}, t = 2,3,4,5; \\
& X_{3,i1} = U_{3i};\ X_{3,it} = \rho X_{3,it-1} + \sigma V_{3,it}, t = 2,3,4,5; \\
& \alpha_i = \overline{X}_{1,i}/3 + \overline{X}_{2,i}/3 + \overline{X}_{3,i}/3 - (1 + \rho + \rho^2 + \rho^3 + \rho^4 + \rho^3\sigma + 2\rho^2\sigma + 3\rho\sigma + 4\sigma)/6 \cdot \left(\sqrt{\frac{2}{\pi}} + 1.5\right) \\
& Y_{it} = \alpha_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + U_{it},\ i = 1, \cdots, n; t = 1,2;\ \ U_{it} \sim \mathcal{N}(0,1).
\end{aligned}
$$

Table 2: Comparison of Pseudo and FE Estimates (DGP2, $\rho = 0.8, \sigma = 0.2$)

Root Mean Square Errors									
n	RMSE of IE			RMSE of FE			RRMSE		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
200	0.205	0.181	0.324	0.214	0.114	0.426	1.046	0.632	1.316
400	0.150	0.127	0.235	0.154	0.076	0.319	1.027	0.595	1.358
800	0.113	0.096	0.212	0.093	0.062	0.214	0.824	0.645	1.006

6 Empirical Illustration

We illustrate our method in two genuine panel data applications. First is the 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 on 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 "calibration" matching 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). Matching 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. Our elongated pseudo panel is a balanced 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 of the tables below.

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 matching method, whereas some effects are negative in the original paper.

Based on our 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 matching removing 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)					Pseudo Panel Matching				
	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
	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
	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
	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)					Pseudo Panel Matching				
	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
	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
	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
	2.788	0.815	1.594	1.668	1.600	12.427	22.388	1.494	11.432	11.383

Table 3: Static and Dynamic homogeneous models (Table 5 and 6 in EHS). POLS: pooled OLS with time FEs; 2FE: two-way fixed effects (in the dynamic models, imposing COMFAC restriction); FD: first differences with time FEs; CCEP: pooled common correlated effects (without and with year dummies); BB: dynamic micropanel estimator by Blundell and Bond (1998).

Heterogenous Models (Static)								
Eberhardt, Helmers, and Strauss (2013)					Pseudo Panel Matching			
	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
	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
	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
	-0.728	2.123	0.445	-0.601	7.458	4.958	7.034	6.938
Heterogeous Models (Dynamic)								
Eberhardt, Helmers, and Strauss (2013)					Pseudo Panel Matching			
	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
	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
	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
	-0.953	3.947	-0.945	-0.964	8.126	9.847	4.985	4.145

Table 4: Heterogeneous models (Table 7 and 8in EHS). MG: Mean group estimator by Pesaran and Smith (1995); CDMG: cross-section demeaned mean group; CMG: Pesaran (2006) CCE mean group version (without and with year dummies).

6.2 Female Labor Force Participation

We illustrate the empirical performance of our methods in the study of female labor force participation as in Fernández-Val (2009). The original panel is from PSID Waves 13-22, with 1461 females over 10 years. Fernández-Val proposes a bias correction method to handle the incidental parameters problem in the nonlinear panel data model. Fernández -Val examined the relationship between fertility and female labor force participation. He employed 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 information based matching method to generate a "pseudo" panel based on the calibration estimation of the aggreagator function. Matching 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 following tables. The results of linear probability model are similar. Our results are also consistent with 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 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 (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 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 matching entire samples.

	Fernández-Val (2009)				Pseudo Panel Matching			
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 5: Static model [Table 10 (Probit Part)] in Fernández-Val (2009). FE denotes uncorrected fixed effects estimator; JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC3 denotes the bias-corrected estimator proposed in Fernández-Val (2009); BC3p denotes the bias-corrected estimator proposed in this paper when the regressors are treated as predetermined.

REFERENCES

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 (2016). Matching on the estimated propensity score. *Econometrica 84*(2), 781–807.

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.

	Fernández-Val (2009)					Pseudo Panel Matching				
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 6: Static model [Table 10 (Logit Part)] in Fernández-Val (2009). C denotes conditional logit estimator.

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.

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. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics 87*(1), 115–143.

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.

	Fernández-Val (2009)				Pseudo Panel Matching			
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 7: Dynamic model [Table 11, Probit and Logit] in Fernández-Val (2009)

Blundell, R., T. MaCurdy, and C. Meghir (2007). Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics. *Handbook of Econometrics 6*, 4667–4775.

Blundell, R. and T. M. Stoker (2005). Heterogeneity and aggregation. *Journal of Economic Literature 43*(2), 347–391.

Blundell, R. and T. M. Stoker (2007). Models of aggregate economic relationships that account for heterogeneity. *Handbook of Econometrics 6*, 4609–4666.

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

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.

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. (1984). Panel data. *Handbook of econometrics 2*, 1247–1318.

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? *Review of Economics and Statistics 95*(2), 436–448.

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 (2017). A similarity based model for ordered categorical data. *Accepted by Econometric Reviews*.

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.

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 Jour-*

nal 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.

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.

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.

Maasoumi, E. and T. Xu (2015). Weights and substitution degree in multidimensional well-being in china. *Journal of Economic Studies 42*(1), 4–19.

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.

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.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics 68*(1), 79–113.

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.

Seawright, J. (2009). Matching for pseudo-panel inference. *Working paper*.

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

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