THE UNIVERSITY OF SYDNEY

The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Consistent Estimation of Linear Regression Models Using Matched Data*

Masayuki Hirukawa
Setsunan University

Artem Prokhorov
University of Sydney

Abstract

Economists often use matched samples, especially when dealing with earnings data where a number of missing observations need to be imputed. In this paper, we demonstrate that the ordinary least squares estimator of the linear regression model using matched samples is inconsistent and has a non-standard convergence rate to its probability limit. If only a few variables are used to impute the missing data then it is possible to correct for the bias. We propose two semi-parametric bias-corrected estimators and explore their asymptotic properties. The estimators have an indirect-inference interpretation and their convergence rates depend on the number of variables used in matching. We can attain the parametric convergence rate if that number is no greater than three. Monte Carlo simulations confirm that the bias correction works very well in such cases.

Keywords: Bias correction; differencing; indirect inference; linear regression; matching estimation; measurement error bias.

JEL Classifcation Codes: C13; C14; C31.

May 2014

BA Working Paper No: BAWP-2014-03
http://sydney.edu.au/business/business_analytics/research/working_papers

Consistent Estimation of Linear Regression Models Using Matched Data*

Masayuki Hirukawa† Artem Prokhorov‡
Setsunan University University of Sydney

May 2014

Abstract

Economists often use matched samples, especially when dealing with earnings data where a number of missing observations need to be imputed. In this paper, we demonstrate that the ordinary least squares estimator of the linear regression model using matched samples is inconsistent and has a nonstandard convergence rate to its probability limit. If only a few variables are used to impute the missing data then it is possible to correct for the bias. We propose two semiparametric bias-corrected estimators and explore their asymptotic properties. The estimators have an indirect-inference interpretation and their convergence rates depend on the number of variables used in matching. We can attain the parametric convergence rate if that number is no greater than three. Monte Carlo simulations confirm that the bias correction works very well in such cases.

Keywords: Bias correction; differencing; indirect inference; linear regression; matching estimation; measurement error bias.

JEL Classification Codes: C13; C14; C31.

---

*The first author gratefully acknowledges financial support from Japan Society of the Promotion of Science (grant number 23530259).

†Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka 572-8508, Japan; phone: (+81)72-839-8095; fax: (+81)72-839-8138; e-mail: hirukawa@econ.setsunan.ac.jp.

‡Discipline of Business Analytics, Business School, University of Sydney, H04-499 Merewether Building, Sydney, NSW 2006, Australia; phone: (+61)2-9351-6584; fax: (+61)2-9351-6409; e-mail: artem.prokhorov@sydney.edu.au.

1 Introduction

Suppose that we are interested in estimating a linear regression model

$$Y = X_1'\beta_1 + X_2'\beta_2 + Z'\gamma + u := W'\theta + u,\ E\left(u|\,W\right) = 0, \tag{1}$$

using a random sample, where $X_1 \in \mathbb{R}^{d_1}$, $X_2 \in \mathbb{R}^{d_2}$ and $Z \in \mathbb{R}^{d_3}$. (The reason for distinguishing between the regressors X_1, X_2 and Z will become clear shortly.) While $d_1 = 0$ is allowed (see, e.g., Example 2 below), the intercept term is assumed to be included as a component of X_2 so that $d_2 \geq 2$ must be the case. When $W = (X_1', X_2', Z')' \in \mathbb{R}^d$, where $d := d_1 + d_2 + d_3$, is exogenous and a single random sample of (Y, W) can be obtained, the ordinary least squares (OLS) estimator of $\theta = (\beta_1', \beta_2', \gamma')'$ is consistent and even Gauss-Markov when the error term u is conditionally homoskedastic.

In reality, however, we often face the problem that (Y, W) cannot be taken from a single data source. It is not uncommon in labor and public economics to collect the variables necessary for regression analysis from more than one sources. Examples include Lusardi (1996), Björklund and Jäntti (1997), Currie and Yelowitz (2000), Dee and Evans (2003), Borjas (2004), and Fujii (2008), to name a few. This is the setting in which we are interested. Specifically, suppose that instead of observing a complete data set (Y, W), we have the following two overlapping subsets of the data, (Y, X_1, Z) and (X_2, Z). That is, some of the regressors are not available in the initial data set, where the initial data set is the data set containing observations on the dependent variable along with a few other regressors. In such a setting, it is natural to construct a matched data set via exploiting the proximity of the common regressor(s) Z across the two samples. A common method of constructing matched samples is the nearest neighbor method (NNM; see, e.g., Abadie and Imbens, 2006, 2012a). Here are a few examples of the setting.

Example 1. (Earnings data) Matching is currently used for imputing missing records of earnings in important economic data sets. For example, the U.S. Current Population Survey (CPS) files use the so called "hot deck imputation" procedure of the Census (see, e.g., Little and Rubin, 2002; Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006), which allocates to nonrespondents the reported earnings of a matched respondent who has similar recorded attributes. The share of imputed values is as high as 30%. The resulting earnings data have been used to uncover much of what is known about the labor market dynamics and outcomes.

Example 2. (Intergenerational income mobility) Let Y denote (the logarithm of) son's earnings, X_2 (the logarithm of) father's earnings and father's individual characteristics, and Z father's education. In this example, X_1 is absent and all the new regressors we can obtain (aside from Z) will come from matching on Z. A complete data set of (Y, W) is unavailable but we can take the two data sets (Y, Z) and (X_2, Z) separately from two Panel Study of Income Dynamics (PSID) waves with a gap of 20 years, for instance, and construct a matched data set of (Y, W) where the matching is based on Z.

Example 3. (Return to schooling) Let Y denote (the logarithm of) earnings, X_1 individual characteristics, X_2 ability measured by test scores, and Z education. Although (Y, X_1, Z) is available in PSID, for instance, it is often the case that (X_2, Z) can be found only in a different, psychometric data set. Again, we must construct a matched data set of (Y, W) with respect to Z. Unlike in Example 2, here we use the matched observations for only a subset of regressors.

This paper demonstrates that the OLS estimation of (1) using NNM-based matching is inconsistent. The source of the inconsistency is a non-vanishing nonparametric bias term, which can be viewed as a measurement error bias stemming from replacing unobservables X_2 with their proxy in the matched data. In this sense, the paper is related to the literature on the classical problem of generated regressors and missing data (see, e.g., Pagan, 1984; Prokhorov and Schmidt, 2009). Moreover, we show that the rate of convergence to the probability limit of OLS depends on the number of matching variables. In particular, the parametric rate is attained only when $d_3 = 1$, i.e. when there is only one matching variable. In line with these findings, the paper proposes two semiparametric bias-corrected estimators. The first estimator is based on the original regression (1), which is in levels, and is designed only for the case with $d_3 = 1$. The second estimator attempts to remedy the curse of dimensionality in d_3 by using first differences in (1). We show that it attains the parametric convergence rate as long as $d_3 \leq 3$. Both estimators can be interpreted as indirect inference estimators (Gouriéroux, Monfort and Renault, 1993; Smith, 1993) in the sense that they can be obtained by taking the probability limit of the OLS estimator from the regression as the "binding" function.

Correspondingly, this paper makes contributions in three important areas. First, we provide new asymptotic results for regressions involving matched data. Such results have been limited to the literature on matching estimators of the average treatment effect (ATE). For example, Abadie and Imbens (2006) show that when there is only one matching covariate, the bias in NNM-based matching estimators of the ATE may be asymptotically ignored; they attain the parametric convergence rate in that case.

Second, the estimation theory we develop provides a guidance on repeated survey sampling when some covariates are found to be completely missing after the initial survey. Our theory suggests (approximately) how many observations should be collected in a follow-up survey and how to estimate the linear regression model of interest consistently using the matched data from it.

Finally, the paper offers an alternative to other estimation methods based on two samples. A number of such methods have been designed within the framework of instrumental variables (IV) or generalized method of moments (GMM) estimation, where we can construct required moments off the two samples individually so no matching is required (e.g., Angrist and Krueger, 1992, 1995; Arellano and Meghir, 1992; Inoue and Solon, 2010; Murtazashvili, Liu and Prokhorov, 2013). This approach is not applicable in the setting of a linear regression where some regressors are missing and two-sample moment based estimation is infeasible.

In this paper we assume that the two samples *jointly* identify the regression models. There are other two-sample estimators, e.g., Imbens and Lancaster (1994) and Hellerstein and Imbens (1999), that cover the cases where the first sample *alone* identifies the models and the second sample is used for efficiency gains. These are not the settings we consider.

The paper is organized as follows. Section 2 shows inconsistency of the OLS estimation of the regression model in (1). Section 3 proposes two bias-corrected estimators and explores their convergence properties. We also discuss a consistent estimation of the covariance matrix in Section 3. Section 4 discusses the results of Monte Carlo simulations, which examine how the bias correction works in finite samples. Section 5 concludes with a few questions for future research. All proofs are given in the Appendix.

The paper adopts the following notational conventions: $\|A\| = \{\mathrm{tr}\,(A'A)\}^{1/2}$ is the Euclidian norm of matrix A; $0_{p\times q}$ signifies the $p \times q$ zero matrix, where the subscript may be suppressed if $q = 1$; and $c\,(> 0)$ denotes a generic constant, which is different from one statement to another.

2 Inconsistency of OLS Estimation Using Matched Samples

In order to explain how a matched sample is constructed, we need more notations. Denote the two random samples by $\mathcal{S}_1$ and $\mathcal{S}_2$. Also let n and m be the sizes of $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. Specifically, the two samples can be expressed as $\mathcal{S}_1 = \mathcal{S}_{1n} = \{(Y_i, X_{1i}, Z_i)\}_{i=1}^n$ and $\mathcal{S}_2 = \mathcal{S}_{2m} = \{(X_{2j}, Z_j)\}_{j=1}^m$. A natural way to match based on Z is by using the NNM, which chooses

$$Z_{j(i)} := \arg \min_{1 \le j \le m} \|Z_j - Z_i\|$$

for each Z_i $(1 \le i \le n)$.

There are many other methods based on various metrics (see, e.g., Smith and Todd, 2005), but we focus on the NNM because it seems to be most prevalent in applied research, especially in the ATE literature. Also, we consider the NNM based on a single match, where matching is done with replacement, and each element of the matching vector Z is assumed to be continuous. So our setting can be viewed as a special case of M matches for the ATE estimation considered, e.g., by Abadie and Imbens (2006) and as a foundation for more complicated methods of kernel-based matching (see, e.g., Busso, DiNardo, McCrary, 2014). Matching with replacement, allowing each unit to be used as a match more than once, seems to be standard in econometric literature, while the inclusion of discrete matching variables with a finite number of support points does not affect the subsequent asymptotic results. Finally, for simplicity, we ignore ties in the NNM, which happen with probability zero as long as Z is continuous.

Applying the NNM we obtain a matched data set $\mathcal{S}_n = \left\{ \left(Y_i, X_{1i}, X_{2j(i)}, Z_i, Z_{j(i)} \right) \right\}_{i=1}^n$, where $X_{2j(i)}$ is the observation paired with $Z_{j(i)}$ in $\mathcal{S}_2$. Throughout, it is assumed that we estimate θ by regressing Y_i on $W_{i,j(i)} := \left(X'_{1i}, X'_{2j(i)}, Z'_{j(i)} \right)'$. Alternatively, we could use Z_i in place of $Z_{j(i)}$. However, both alternatives are first-order asymptotically equivalent, after the bias correction, and so we concentrate exclusively on the former case.

The OLS estimator

$$\hat{\theta}_{OLS} := \hat{Q}_W^{-1}\hat{R}_W := \left(\frac{1}{n}\sum_{i=1}^{n} W_{i,j(i)}W_{i,j(i)}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} W_{i,j(i)}Y_i$$

is referred to as the *matched-sample OLS* (MSOLS) estimator hereafter. It will be shown shortly that the MSOLS estimator is inconsistent. Demonstrating this result and deriving the bias-corrected, consistent estimators of θ require the following assumptions.

Assumption 1. Two random samples $(\mathcal{S}_1, \mathcal{S}_2) = (\mathcal{S}_{1n}, \mathcal{S}_{2m})$ are drawn independently from the joint distribution of (Y, W) with finite fourth-order moments, where the two sample sizes satisfy $n/m \to \kappa \in (0, \infty)$ as $n, m \to \infty$.

Assumption 2. The matching variable Z is continuously distributed with a convex and compact support $\mathbb{Z}$, with the density bounded and bounded away from zero on its support.

Assumption 3.

(i) The regression error u satisfies $E(u|W) = 0$ and $\sigma_u^2(W) := E(u^2|W) \in (0, \infty)$.

(ii) Let $g(Z) := \left[\begin{array}{cc} g_1(Z)' & g_2(Z)' \end{array}\right]' := \left[\begin{array}{cc} E(X_1|Z)' & E(X_2|Z)' \end{array}\right]'$ and let $\eta := \left[\begin{array}{cc} \eta_1' & \eta_2' \end{array}\right]' := \left[\begin{array}{cc} X_1' - g_1(Z)' & X_2' - g_2(Z)' \end{array}\right]'$. Then, $\Sigma_1 := E(\eta_1\eta_1') > 0$, $\Sigma_2 := E(\eta_2\eta_2') \geq 0$ with $\text{rank}(\Sigma_2) = d_2 - 1$, $E(\eta_1\eta_2') = 0_{d_1 \times d_2}$, and $g(\cdot)$ is first-order Lipschitz continuous on $\mathbb{Z}$.

These regularity conditions are largely inspired by those in the literature on semiparametric, partial linear regression models (e.g., Robinson, 1988; Yatchew, 1997), matching estimators for the ATE (e.g., Abadie and Imbens, 2006), and regression estimation based on two samples (e.g., Angrist and Krueger, 1992; Inoue and Solon, 2010). Assumption 1 refers to the same divergence rate of the two sample sizes. This condition can be commonly found in the literature on two-sample regression estimation. Assumption 2 plays a key role in controlling the order of magnitude in the *matching discrepancy* (Abadie and Imbens, 2006), the definition of which can be found in the Appendix. The rank condition in Assumption 3(ii) comes from the fact that the row and column of Σ_2 corresponding to the intercept are identically zero. While uncorrelatedness between η_1 and η_2 in this assumption appears to be restrictive, the condition simplifies subsequent analysis considerably.

We start our asymptotic analysis from rewriting Y_i in a 'partial linear'-like format. A straightforward calculation yields

$$Y_i := W'_{i,j(i)}\theta + \lambda_{i,j(i)} + \epsilon_{i,j(i)},\ i = 1, \dots, n, \tag{2}$$

where

$$\lambda_{i,j(i)} = \lambda\left(Z_i, Z_{j(i)}\right) = \left\{g_2\left(Z_i\right) - g_2\left(Z_{j(i)}\right)\right\}' \beta_2 + \left(Z_i - Z_{j(i)}\right)' \gamma, \text{ and}$$
$$\epsilon_{i,j(i)} = u_i + \left(\eta_{2i} - \eta_{2j(i)}\right)' \beta_2.$$

The reason why this is not exactly a partial linear model is that there is a common regressor $Z_{j(i)}$ included in $W_{i,j(i)}$ and $\lambda_{i,j(i)}$. In this formulation, $W_{i,j(i)}$ is employed as the regressor of the fully parametric part $W'_{i,j(i)}\theta$, whereas the semiparametric part $\lambda_{i,j(i)}$ could be viewed as an analog to the summand for the conditional bias in the matching estimator investigated in Abadie and Imbens (2006). A key difference from the partial linear regression models studied in Robinson (1988) and Yatchew (1997) is that the matched regressor $X_{2j(i)}$ is endogenous, i.e., $X_{2j(i)}$ and the composite error $\epsilon_{i,j(i)}$ are correlated. The theorem below is established for the model in (2); it provides the probability limit of $\hat{\theta}_{OLS}$ with the rate of convergence.

Theorem 1. *If Assumptions 1-3 hold and* $Q_W := E\left(W_{i,j(i)}W'_{i,j(i)}\right) > 0$, *then* $\hat{\theta}_{OLS} = Q_W^{-1}P_W\theta + O_p\left(n^{-\min\{1/2,1/d_3\}}\right)$, *where* $P_W := Q_W - \Sigma$ *and* Σ *is a* $d \times d$ *block-diagonal matrix of the form* $\Sigma := \text{diag}\{0_{d_1\times d_1}, \Sigma_2, 0_{d_3\times d_3}\}$.

The theorem states that MSOLS is inconsistent in general. The term Σ in P_W, which is the source of the inconsistency, is generated by misspecifying the regression of Y_i on W_i as the one of Y_i on $W_{i,j(i)}$, or equivalently, employing $X_{2j(i)}$ as a proxy of the latent variable X_{2i}. Therefore, the non-vanishing bias in MSOLS can be thought of as a measurement error bias. We can also find from a straightforward calculation that the OLS estimator of β_2, which is the coefficient vector on the matched regressor X_2, is biased toward zero in the limit.

A quick inspection also reveals that $\hat{\theta}_{OLS}$ would be consistent if either (i) $\beta_2 = 0$, i.e. X_2 were irrelevant in the correctly specified model; or (ii) $\Sigma_2 = 0$, i.e. X_2 were expressed as a *nonlinear* deterministic function of Z. The nonlinearity in $g_2(\cdot)$ is important because perfect multicollinearity would occur if the function were linear but also because the nonlinearity plays a key role in our bias-corrected estimation of θ (see Remark 1 in Section 3). However, even if either one of the above conditions were true, $\hat{\theta}_{OLS}$ might not be $\sqrt{n}$-consistent. Clearly, there exists a *curse of dimensionality* with respect to the matching variable Z. The proof of Theorem 1 in the Appendix suggests that when $d_3 = 1$, $\sqrt{n}\left(\hat{\theta}_{OLS} - Q_W^{-1}P_W\theta\right)$ has a normal limit. When $d_3 = 2$, $\hat{\theta}_{OLS}$ is still $\sqrt{n}$-consistent, but we could only demonstrate asymptotic normality of $\hat{\theta}_{OLS}$ after subtracting the bias term due to the matching discrepancy, i.e. the best we can do in this case is to apply the central limit theorem (CLT) to $\sqrt{n}\left(\hat{\theta}_{OLS} - Q_W^{-1}P_W\theta - B_{OLS2}\right)$. These limiting distributions would reduce to the usual one of OLS if a complete data set of (Y, W) were available. When $d_3 \geq 3$, the convergence rate of $\hat{\theta}_{OLS}$ is slower than the parametric one, and it becomes slower as d_3 increases.

3 Bias-Corrected Estimation of the Parameter

3.1 An Overview

This section develops a bias-corrected, consistent estimator of θ. As suggested by the proof of Theorem 1 in the Appendix, inconsistency of MSOLS comes from the fact that $\hat{Q}_W \xrightarrow{p} Q_W$ whereas $\hat{R}_W \xrightarrow{p} P_W\theta = (Q_W - \Sigma)\,\theta$. Therefore, the non-vanishing bias in MSOLS can be eliminated if either

(a) the denominator $\hat{Q}_W$ is replaced by a consistent estimator of P_W with the numerator $\hat{R}_W$ left unchanged; or

(b) an extra term consistent for $\Sigma\theta$ is added to $\hat{R}_W$ with $\hat{Q}_W$ held as it is.

Bias correction in each strategy is semiparametric in that a consistent estimate of Σ_2 (covariance matrix of the nonparametric regression error η_2) is required. Moreover, because implementing (b) must result in a two-step estimation with an initial consistent estimate of θ plugged in, we first explore strategy (a). As discussed shortly, this idea can be interpreted as a variant of indirect inference (II) by Gouriéroux, Monfort and Renault (1993) and Smith (1993). On the other hand, strategy (b) is reminiscent of the fully-modified (FM) least squares estimation for cointegrating regressions by Phillips and Hansen (1990), and it is investigated at the end of this section.

3.2 MSOLS-Based Bias Correction

To obtain a $\sqrt{n}$-consistent and asymptotically normal estimator of θ based on strategy (a), we assume that $d_3 = 1$. The estimator can be interpreted as an II estimator. Take the probability limit of $\hat{\theta}_{OLS}$ as the binding function $b(\theta)$, i.e. $b(\theta) = Q_W^{-1} P_W \theta$.[1] Provided that P_W^{-1} exists, the II estimator can be built on the inverse mapping of $\hat{\theta}_{OLS} = b(\theta)$, i.e. $\theta = P_W^{-1} Q_W \hat{\theta}_{OLS}$. The interpretation then follows from replacing

[1] Typically the binding function is unknown, and it must be approximated via simulations. However, when the function has a closed form, there is no need for simulations; see Carrasco and Florens (2002) for another example.

P_W with its $\sqrt{n}$-consistent estimator $\hat{P}_W$ and regarding $\hat{R}_W$ as a 'sample analog' to $Q_W \hat{\theta}_{OLS}$. Accordingly, we call this estimation method *the matched-sample indirect inference* (MSII) estimation hereafter. We also refer to the estimator

$$\hat{\theta}_{II-L} = \hat{P}_W^{-1} \hat{R}_W$$

as the MSII-L estimator. The last letter "L" stands for the level regression (2), reflecting that later MSII is also applied to the first difference of (2).

Our remaining task is to deliver a consistent estimator of P_W. Obviously, $\hat{Q}_W$ is a natural estimator of Q_W. Furthermore, it turns out that when estimating $\Sigma = \text{diag}\left\{0_{d_1 \times d_1}, \Sigma_2, 0_{d_3 \times d_3}\right\}$, we can do without a nonparametric estimation of $g_2(\cdot)$. Assume that $\mathcal{S}_2$ is reordered by the ordering rule in Lemma A2 in the Appendix. Then, Σ_2 can be consistently estimated by

$$\hat{\Sigma}_2 = \frac{1}{2(m-1)} \sum_{j=2}^{m} \Delta X_{2j} \Delta X_{2j}', \tag{3}$$

where $\Delta X_{2j} := X_{2j} - X_{2j-1}$. This is known as the difference-based variance estimator suggested by Rice (1984), and Lemma A2 implies that $\hat{\Sigma}_2 = \Sigma_2 + O_p\left(m^{-1/2}\right)$ as long as $d_3 \leq 3$. In the end, the estimator of P_W is given by $\hat{P}_W := \hat{Q}_W - \hat{\Sigma} = \hat{Q}_W - \text{diag}\left\{0_{d_1 \times d_1}, \hat{\Sigma}_2, 0_{d_3 \times d_3}\right\}$.

The next theorem establishes $\sqrt{n}$-consistency of $\hat{\theta}_{II-L}$ and derives its limiting distribution.

Theorem 2. *If Assumptions 1-3 hold, $d_3 = 1$ and P_W^{-1} exists, then $\hat{\theta}_{II-L} \stackrel{p}{\to} \theta$ and $\sqrt{n}\left(\hat{\theta}_{II-L} - \theta\right) \stackrel{d}{\to} N(0, V_W) := N\left(0, P_W^{-1}\Omega_W P_W^{-1}\right)$, where*

$$
\begin{aligned}
\Omega_W &= \Omega_{W,11} + \sqrt{\kappa}\left(\Omega_{W,12} + \Omega_{W,12}'\right) + \kappa\Omega_{W,22}, \\
\Omega_{W,11} &= E\left(\phi_{i,j(i)}\phi_{i,j(i)}'\right), \\
\Omega_{W,12} &= \left[\begin{array}{ccc} 0_{d\times d_1} & E\left(\phi_{i,j(i)}\omega_{j(i)}'\right) & 0_{d\times d_3} \end{array}\right], \\
\Omega_{W,22} &= \text{diag}\left\{0_{d_1\times d_1}, E\left(\psi_j\psi_{j-1}'\right) + E\left(\psi_j\psi_j'\right) + E\left(\psi_j\psi_{j+1}'\right), 0_{d_3\times d_3}\right\}, \\
\phi_{i,j(i)} &= W_{i,j(i)}\epsilon_{i,j(i)} + \Sigma\theta, \\
\omega_{j(i)} &= \left(\eta_{2j(i)}\eta_{2j(i)}' - \Sigma_2\right)\beta_2, \text{ and} \\
\psi_j &= \left(\frac{\Delta\eta_{2j}\Delta\eta_{2j}'}{2} - \Sigma_2\right)\beta_2.
\end{aligned}
$$

Remark 1. There are two important observations here. First, some nonlinearity in all elements of $g_2(\cdot)$ other than the intercept is necessary for a non-singular P_W and thus for identification of θ. To see why, observe that the lower-right block of P_W collapses to

$$
\left[\begin{array}{cc} E\left(X_{2j(i)}X_{2j(i)}'\right) - \Sigma_2 & E\left(X_{2j(i)}Z_{j(i)}'\right) \\ E\left(Z_{j(i)}X_{2j(i)}'\right) & E\left(Z_{j(i)}Z_{j(i)}'\right) \end{array}\right] = \left[\begin{array}{cc} E\left\{g_2(Z)\,g_2(Z)'\right\} & E\left\{g_2(Z)\,Z'\right\} \\ E\left\{Zg_2(Z)'\right\} & E(ZZ') \end{array}\right],
$$

which becomes singular if one or more elements of $g_2(\cdot)$ other than the intercept are linear. Second, Theorem 2 suggests that in the special case where $n = o(m)$, Ω_W reduces to $\Omega_{W,11} = Var\left(\phi_{i,j(i)}\right)$.

3.3 Consistent Estimation for Two or More Matching Variables

While MSII-L yields a consistent estimate of θ, its apparent deficiency is that it can be applied only for the case with a single matching variable. The curse of dimensionality in the NNM can be commonly observed in other applications. With regards to the ATE estimation, Abadie and Imbens (2006, Corollary 1), for instance, show that the matching discrepancy bias can be safely ignored only when matching is done on a single variable.

To provide a remedy for this issue, we follow the strategy in Yatchew (1997). Assume that $\mathcal{S}_1$ and $\mathcal{S}_2$ are both reordered with respect to Z_i and Z_j, respectively, by the ordering rule in Lemma A2 in the Appendix. Then, taking the first-order difference of regression (2) yields

$$\Delta Y_i := \Delta X'_{i,j(i)}\beta + \Delta\mu_{i,j(i)} + \Delta\epsilon_{i,j(i)},\ i = 2, \ldots, n, \tag{4}$$

where $X_{i,j(i)} = \left(X'_{1i}, X'_{2\backslash 0 j(i)}\right)'$, $\beta = \left(\beta'_1, \beta'_{2\backslash 0}\right)'$, $\Delta\mu_{i,j(i)} = \Delta Z'_{j(i)}\gamma + \Delta\lambda_{i,j(i)} = \left\{\Delta g_{2\backslash 0}(Z_i) - \Delta g_{2\backslash 0}\left(Z_{j(i)}\right)\right\}' \beta_{2\backslash 0} + \Delta Z'_i\gamma$, and the subscript "$\bullet\backslash 0$" denotes "with the intercept excluded". Note that both β_0 (the intercept) and γ (the coefficient on the matching variable Z) are not identified from the first-difference regression (4). We can estimate these parameters consistently from the level regression (2) after obtaining a consistent estimate of β.

As in the MSII-L estimation, we start our analysis by deriving the first-order bias of the OLS estimator for this regression

$$\hat{\beta}_{FD} := \hat{Q}_{\Delta X}^{-1}\hat{R}_{\Delta X} := \left(\frac{1}{n-1}\sum_{i=2}^{n}\Delta X_{i,j(i)}\Delta X'_{i,j(i)}\right)^{-1}\frac{1}{n-1}\sum_{i=2}^{n}\Delta X_{i,j(i)}\Delta Y_i,$$

where the abbreviation "FD" stands for the first-difference. The theorem below provides the probability limit of $\hat{\beta}_{FD}$ and its convergence rate.

Theorem 3. *If Assumptions 1-3 hold and* $Q_{\Delta X} := E\left(\Delta X_{i,j(i)}\Delta X'_{i,j(i)}\right) > 0$*, then* $\hat{\beta}_{FD} = Q_{\Delta X}^{-1}P_{\Delta X}\theta + O_p\left(n^{-\min\{1/2, 2(1-\delta)/d_3\}}\right)$*, where* $P_{\Delta X} := Q_{\Delta X} - 2\Sigma_{FD}$*,* $\Sigma_{FD} = \operatorname{diag}\left\{0_{d_1\times d_1}, \Sigma_{2\backslash 0}\right\}$*,* $\Sigma_{2\backslash 0}$ *is the* $(d_2-1)\times(d_2-1)$ *matrix that can be obtained by dropping the row and column corresponding to the intercept from* Σ_2*, and* $\delta\,(>0)$ *is a constant arbitrarily close to* 0*.*

Theorem 3 suggests that a $\sqrt{n}$-consistent estimator of β based on regression (4) can be obtained if $d_3 \leq 3$, i.e. if the number of (continuous) matching variables does not exceed three. Then, taking the binding function $b(\beta) = Q_{\Delta X}^{-1} P_{\Delta X} \beta$ implies that the MSII-FD estimator

$$\hat{\beta}_{II-FD} = \hat{P}_{\Delta X}^{-1} \hat{R}_{\Delta X}$$

is a $\sqrt{n}$-consistent and asymptotically normal estimator of β corresponding to strategy (a), where $\hat{P}_{\Delta X} := \hat{Q}_{\Delta X} - 2\hat{\Sigma}_{FD} = \hat{Q}_{\Delta W} - 2\,\mathrm{diag}\left\{0_{d_1 \times d_1}, \hat{\Sigma}_{2\backslash 0}\right\}$, and $\hat{\Sigma}_{2\backslash 0} := \{2(m-1)\}^{-1} \sum_{j=2}^{m} \Delta X_{2\backslash 0 j} \Delta X_{2\backslash 0 j}'$. The next theorem presents the convergence results for $\hat{\beta}_{II-FD}$. The proof is found to be a minor modification of the one for Theorem 2, and thus it is omitted.

Theorem 4. *If Assumptions 1-3 hold, $d_3 \leq 3$ and $P_{\Delta X}^{-1}$ exists, then $\hat{\beta}_{II-FD} \xrightarrow{p} \beta$ and $\sqrt{n}\left(\hat{\beta}_{II-FD} - \beta\right) \xrightarrow{d} N(0, V_{\Delta X}) := N\left(0, P_{\Delta X}^{-1} \Omega_{\Delta X} P_{\Delta X}^{-1}\right)$. In particular, the analytical expression of $\Omega_{\Delta X}$ is $\Omega_{\Delta X} = \Omega_{\Delta X,11} + 2\sqrt{\kappa}\left(\Omega_{\Delta X,12} + \Omega_{\Delta X,12}'\right) + 4\kappa \Omega_{\Delta X,22}$, where*

$$\begin{aligned}
\Omega_{\Delta X,11} &= E\left(\xi_{i,j(i)} \xi_{i-1,j(i-1)}'\right) + E\left(\xi_{i,j(i)} \xi_{i,j(i)}'\right) + E\left(\xi_{i,j(i)} \xi_{i+1,j(i+1)}'\right), \\
\Omega_{\Delta X,12} &= \left[\begin{array}{cc} 0_{(d_1+d_2-1)\times d_1} & E\left(\xi_{i,j(i)} \zeta_{j(i-1)}'\right) + E\left(\xi_{i,j(i)} \zeta_{j(i)}'\right) + E\left(\xi_{i,j(i)} \zeta_{j(i+1)}'\right) \end{array}\right], \\
\Omega_{\Delta X,22} &= \mathrm{diag}\left\{0_{d_1 \times d_1}, E\left(\zeta_j \zeta_{j-1}'\right) + E\left(\zeta_j \zeta_j'\right) + E\left(\zeta_j \zeta_{j+1}'\right)\right\}, \\
\xi_{i,j(i)} &= \Delta \eta_{i,j(i)} \Delta \epsilon_{i,j(i)} + 2\Sigma_{FD}\beta, \ \text{and} \\
\zeta_j &= \left(\frac{\Delta \eta_{2\backslash 0 j} \Delta \eta_{2\backslash 0 j}'}{2} - \Sigma_{2\backslash 0}\right) \beta_{2\backslash 0}.
\end{aligned}$$

Remark 2. To recover the unidentified parameter $\alpha := (\beta_0, \gamma)'$, we exploit the level regression (2) in the second step. Let $\tilde{Y}_i := Y_i - X_{i,j(i)}' \hat{\beta}_{II-FD}$ and $\tilde{Z}_{j(i)} := \left(1, Z_{j(i)}'\right)'$. Because $E\left(\tilde{Z}_{j(i)} \epsilon_{i,j(i)}\right) = 0$, α can be estimated by OLS from the regression of $\tilde{Y}_i$ on $\tilde{Z}_{j(i)}$. However, by a similar argument to the one used in the proof of Theorem 1, the OLS estimator

$$\hat{\alpha}_{II-FD} = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_{j(i)} \tilde{Z}_{j(i)}'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_{j(i)} \tilde{Y}_i$$

admits the expansion $\hat{\alpha}_{II-FD} = \alpha + O_p\left(n^{-\min\{1/2, 1/d_3\}}\right)$. As a consequence, $\hat{\alpha}_{II-FD}$ is $\sqrt{n}$-consistent and asymptotically normal only when $d_3 = 1$. However, there is little incentive to adopt the two-step estimation in this case, because MSII-L yields a consistent estimate of the full parameter θ in one step. When $d_3 = 2$, $\hat{\alpha}_{II-FD}$ is still $\sqrt{n}$-consistent, but its limiting distribution could be evaluated after the $O_p\left(n^{-1/2}\right)$ bias term due to the matching discrepancy is subtracted. When $d_3 = 3$, the convergence rate of $\hat{\alpha}_{II-FD}$ is a nonparametric rate of $n^{1/3}$. These complications could be viewed as the potential price to pay for higher-dimensional matching.

3.4 Covariance Estimation

Covariance estimation is essential for inference. First, consider a consistent estimator of $V_W = P_W^{-1}\Omega_W P_W^{-1}$, the asymptotic variance of $\sqrt{n}\left(\hat{\theta}_{II-L} - \theta\right)$. Because $\hat{P}_W$ is consistent for P_W, it suffices to deliver a consistent estimator of $\Omega_W = \Omega_{W,11} + \sqrt{\kappa}\left(\Omega_{W,12} + \Omega_{W,12}'\right) + \kappa\Omega_{W,22}$. Let the MSII-L residual be $\hat{\epsilon}_{i,j(i)} := Y_i - W_{i,j(i)}'\hat{\theta}_{II-L}$. Also denote the MSII-L estimator of β_2 as $\hat{\beta}_{2,II-L}$. Moreover, define $\hat{\phi}_{i,j(i)} := W_{i,j(i)}\hat{\epsilon}_{i,j(i)} + \hat{\Sigma}\hat{\theta}_{II-L}$ and $\hat{\psi}_j := \left\{\left(\Delta X_{2j}\Delta X_{2j}'/2\right) - \hat{\Sigma}_2\right\}\hat{\beta}_{2,II-L}$. Then, a natural estimator of $\Omega_{W,11}$ is

$$\hat{\Omega}_{W,11} = \frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_{i,j(i)}\hat{\phi}_{i,j(i)}'.$$

In addition, $E\left(\psi_j\psi_{j-k}'\right)$, $k = \pm 1, 0$, can be consistently estimated as

$$\begin{aligned}\hat{E}\left(\psi_j\psi_{j-k}'\right) &= \frac{1}{m-1}\sum_{j=\max\{2,2+k\}}^{\min\{m,m+k\}}\hat{\psi}_j\hat{\psi}_{j-k}' \\ &= \frac{1}{m-1}\sum_{j=\max\{2,2+k\}}^{\min\{m,m+k\}}\left(\frac{\Delta\eta_{2j}\Delta\eta_{2j}'}{2} - \hat{\Sigma}_2\right)\hat{\beta}_{2,II-L}\hat{\beta}_{2,II-L}'\left(\frac{\Delta\eta_{2j-k}\Delta\eta_{2j-k}'}{2} - \hat{\Sigma}_2\right) \\ &\quad + o_p\left(m^{-1/2}\right),\end{aligned}$$

where the second equality holds for $d_3 \leq 3$. Hence, a natural estimator of $\Omega_{W,22}$ is given by

$$\hat{\Omega}_{W,22} = \operatorname{diag}\left\{0_{d_1\times d_1}, \hat{E}\left(\psi_j\psi_{j-1}'\right) + \hat{E}\left(\psi_j\psi_j'\right) + \hat{E}\left(\psi_j\psi_{j+1}'\right), 0_{d_3\times d_3}\right\}.$$

Moreover, it follows from the proof of Theorem 2 that $E\left(\phi_{i,j(i)}\omega'_{j(i)}\right) = 2E\left(\phi_{i,j(i)}\psi'_{j(i)}\right)$. Then, a consistent estimator of $E\left(\phi_{i,j(i)}\omega'_{j(i)}\right)$ takes the form of

$$\begin{aligned}
\hat{E}\left(\phi_{i,j(i)}\omega'_{j(i)}\right) &= 2\hat{E}\left(\phi_{i,j(i)}\psi'_{j(i)}\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}\hat{\phi}_{i,j(i)}\hat{\psi}'_{j(i)} \\
&= \frac{2}{n}\sum_{i=1}^{n}\left(W_{i,j(i)}\hat{\epsilon}_{i,j(i)} + \hat{\Sigma}\hat{\theta}_{II-L}\right)\hat{\beta}'_{2,II-L}\left(\frac{\Delta\eta_{2j}\Delta\eta'_{2j}}{2} - \hat{\Sigma}_2\right) + o_p\left(n^{-1/2}\right)
\end{aligned}$$

as before. Therefore, $\Omega_{W,12}$ can be estimated as

$$\hat{\Omega}_{W,12} = \left[\begin{array}{ccc} 0_{d\times d_1} & \hat{E}\left(\phi_{i,j(i)}\omega'_{j(i)}\right) & 0_{d\times d_3} \end{array}\right].$$

Since $n/m = \kappa + o(1)$, we finally obtain an estimator of V_W as

$$\hat{V}_W = \hat{P}_W^{-1}\hat{\Omega}_W\hat{P}_W^{-1} = \hat{P}_W^{-1}\left\{\hat{\Omega}_{W,11} + \sqrt{\frac{n}{m}}\left(\hat{\Omega}_{W,12} + \hat{\Omega}'_{W,12}\right) + \left(\frac{n}{m}\right)\hat{\Omega}_{W,22}\right\}\hat{P}_W^{-1}.$$

Second, consider a consistent estimator of $V_{\Delta X} = P_{\Delta X}^{-1}\Omega_{\Delta X}P_{\Delta X}^{-1}$, the asymptotic variance of $\sqrt{n}\left(\hat{\beta}_{II-FD} - \beta\right)$. As before, we only need to consider a consistent estimator of $\Omega_{\Delta X} = \Omega_{\Delta X,11} + 2\sqrt{\kappa}\left(\Omega_{\Delta X,12} + \Omega'_{\Delta X,12}\right) + 4\kappa\Omega_{\Delta X,22}$. Let the MSII-FD residual be $\widehat{\Delta\epsilon}_{i,j(i)} := \Delta Y_i - \Delta X'_{i,j(i)}\hat{\beta}_{II-FD}$. Also denote the MSII-FD estimator of $\beta_{2\backslash 0}$ as $\hat{\beta}_{2\backslash 0,II-FD}$. Furthermore, define $\hat{\xi}_{i,j(i)} := \Delta X_{i,j(i)}\widehat{\Delta\epsilon}_{i,j(i)} + 2\hat{\Sigma}_{FD}\hat{\beta}_{II-FD}$ and $\hat{\zeta}_j = \left\{\left(\Delta X_{2\backslash 0j}\Delta X'_{2\backslash 0j}/2\right) - \hat{\Sigma}_{2\backslash 0}\right\}\hat{\beta}_{2\backslash 0,II-FD}$. Then, natural estimators of $\Omega_{\Delta X,11}$, $\Omega_{\Delta X,12}$ and $\Omega_{\Delta X,22}$ are

$$\begin{aligned}
\hat{\Omega}_{\Delta X,11} &= \hat{E}\left(\xi_{i,j(i)}\xi'_{i-1,j(i-1)}\right) + \hat{E}\left(\xi_{i,j(i)}\xi'_{i,j(i)}\right) + \hat{E}\left(\xi_{i,j(i)}\xi'_{i+1,j(i+1)}\right), \\
\hat{\Omega}_{\Delta X,12} &= \left[\begin{array}{cc} 0_{(d_1+d_2-1)\times d_1} & \hat{E}\left(\xi_{i,j(i)}\zeta'_{j(i-1)}\right) + \hat{E}\left(\xi_{i,j(i)}\zeta'_{j(i)}\right) + \hat{E}\left(\xi_{i,j(i)}\zeta'_{j(i+1)}\right) \end{array}\right], \text{ and} \\
\hat{\Omega}_{\Delta X,22} &= \operatorname{diag}\left\{0_{d_1\times d_1}, \hat{E}\left(\zeta_j\zeta'_{j-1}\right) + \hat{E}\left(\zeta_j\zeta'_j\right) + \hat{E}\left(\zeta_j\zeta'_{j+1}\right)\right\},
\end{aligned}$$

respectively, where, for $k = \pm 1, 0$,

$$\hat{E}\left(\xi_{i,j(i)}\xi'_{i-1,j(i-k)}\right) = \frac{1}{n-1}\sum_{i=\max\{2,2+k\}}^{\min\{n,n+k\}} \hat{\xi}_{i,j(i)}\hat{\xi}'_{i-k,j(i-k)},$$

$$\hat{E}\left(\xi_{i,j(i)}\zeta'_{j(i-k)}\right) = \frac{1}{n-1}\sum_{i=\max\{2,2+k\}}^{\min\{n,n+k\}} \hat{\xi}'_{i,j(i)}\hat{\zeta}'_{j(i-k)}, \text{ and}$$

$$\hat{E}\left(\zeta_j\zeta'_{j-k}\right) = \frac{1}{m-1}\sum_{j=\max\{2,2+k\}}^{\min\{m,m+k\}} \hat{\zeta}_j\hat{\zeta}'_{j-k}.$$

In the end, $V_{\Delta X}$ can be estimated as

$$\hat{V}_{\Delta X} = \hat{P}_{\Delta X}^{-1}\hat{\Omega}_{\Delta X}\hat{P}_{\Delta X}^{-1} = \hat{P}_{\Delta X}^{-1}\left\{\hat{\Omega}_{\Delta X,11} + 2\sqrt{\frac{n}{m}}\left(\hat{\Omega}_{\Delta X,12} + \hat{\Omega}'_{\Delta X,12}\right) + 4\left(\frac{n}{m}\right)\hat{\Omega}_{\Delta X,22}\right\}\hat{P}_{\Delta X}^{-1}.$$

The following proposition refers to consistency of the covariance estimators. This proposition can be established by the techniques employed for the proofs of Theorems 1-4, and thus the proof is omitted.

Proposition 1. *Suppose that Assumptions 1-3 hold. If* $d_3 = 1$ *and* P_W^{-1} *exists, then* $\hat{V}_W \xrightarrow{p} V_W$. *If* $d_3 \leq 3$ *and* $P_{\Delta X}^{-1}$ *exists, then* $\hat{V}_{\Delta X} \xrightarrow{p} V_{\Delta X}$.

3.5 FM-type versus II-type Estimators

So far consistent estimation of the parameter has been explored based on strategy (a). Here, following strategy (b), we derive the asymptotic properties of the two-step FM-type estimators. Specifically, given $\hat{\theta}_{II-L}$ as the first-step estimate, the FM-type estimator for the level regression in (2) is defined as

$$\hat{\theta}_{FM-L} = \hat{Q}_W^{-1}\left(\hat{R}_W + \hat{\Sigma}\hat{\theta}_{II-L}\right).$$

Likewise, given $\hat{\beta}_{II-FD}$ as the first-step estimate, the FM-type estimator for the first-differenced regression (4) is also defined as

$$\hat{\beta}_{FM-FD} = \hat{Q}_{\Delta X}^{-1}\left(\hat{R}_{\Delta X} + 2\hat{\Sigma}_{FD}\hat{\beta}_{II-FD}\right).$$

Each of the two estimators is consistent by construction. Moreover, they are first-order asymptotically equivalent to their corresponding first-step estimates, although they may differ numerically in finite samples. The proposition below summarizes these claims. Because our Monte Carlo study also indicates that there is very little difference in finite sample performance between FM- and II-estimators for a given regression model, we do not pursue FM-type estimation any further.

Proposition 2. *Suppose that Assumptions 1-3 hold. If $d_3 = 1$ and P_W^{-1} and Q_W^{-1} both exist, then $\hat{\theta}_{FM-L} \xrightarrow{p} \theta$ and $\sqrt{n}\left(\hat{\theta}_{FM-L} - \theta\right) \xrightarrow{d} N\left(0, V_W\right)$. If $d_3 \leq 3$ and $P_{\Delta X}^{-1}$ and $Q_{\Delta X}^{-1}$ both exist, then $\hat{\beta}_{FM-FD} \xrightarrow{p} \beta$ and $\sqrt{n}\left(\hat{\beta}_{FM-FD} - \beta\right) \xrightarrow{d} N\left(0, V_{\Delta X}\right)$.*

4 Finite Sample Performance

4.1 Monte Carlo Setup

We conduct Monte Carlo simulations to examine finite sample properties of the MSII estimation. Consider the regression model

$$Y = \beta_0 + \beta_1 X + \gamma Z + u, \tag{5}$$

where the two samples, namely, $\mathcal{S}_1 = \{(Y_i, Z_i)\}_{i=1}^n$ and $\mathcal{S}_2 = \{(X_j, Z_j)\}_{j=1}^m$, are assumed to be observable. The simulation setup corresponds to the setting of Example 2 in the Introduction, with the regressors $(1, X)$ denoted by X_2 in (1). The data are generated in the following manner. First, we draw two independent samples $\{Z_i\}_{i=1}^n, \{Z_j\}_{j=1}^m \overset{iid}{\sim} U[-2, 2]$. Second, given $\{Z_i\}_{i=1}^n$ and $\{Z_j\}_{j=1}^m$, $\{X_i\}_{i=1}^n$ and $\{X_j\}_{j=1}^m$ are generated by $X = g(Z) + \eta$, where $\eta \overset{iid}{\sim} N(0, 1)$ and $g(z)$ takes one of the following six functional forms:

$$\begin{aligned}
&\text{A} : g(z) = z^2 \\
&\text{B} : g(z) = z^3 \\
&\text{C} : g(z) = \exp(z) \\
&\text{D} : g(z) = z + 2\sin(\pi z) \\
&\text{E} : g(z) = z + (5/\tau)\,\phi(z/\tau), \tau = 0.9 \\
&\text{F} : g(z) = z + (5/\tau)\,\phi(z/\tau), \tau = 0.3
\end{aligned}$$

with $\phi(\cdot)$ being the pdf of $N(0,1)$. Model A is convex, and each of Models B and C is monotone increasing. While these three functions are purely nonlinear, the remaining three models may be thought of as 'intermediate' cases between linear and nonlinear functions in that each of these is constructed as a linear combination of linear and nonlinear functions. In particular, Model D is specified as linear with a cycle. Models E and F, inspired by the Monte Carlo design of Horowitz and Spokoiny (2001), can be viewed as linear with a 'spike'. Third, using $\{(X_i, Z_i)\}_{i=1}^n$, we generate $\{Y_i\}_{i=1}^n$ based on the regression model (5), where $\beta_0 = \beta_1 = \gamma = 1$ and $u \overset{iid}{\sim} N(0,1)$. This procedure provides us with two observable samples $\mathcal{S}_1 = \{(Y_i, Z_i)\}_{i=1}^n$ and $\mathcal{S}_2 = \{(X_j, Z_j)\}_{j=1}^m$, as well as an unobservable complete sample $\mathcal{S}^* = \{(Y_i, X_i, Z_i)\}_{i=1}^n$. Finally, the matched sample $\mathcal{S} = \left\{\left(Y_i, X_{j(i)}, Z_i, Z_{j(i)}\right)\right\}_{i=1}^n$ can be constructed via the NNM with respect to Z.

With regards to sample sizes, for each of $n \in \{500, 1000, 2000\}$, m is chosen as one of $m \in \{n/2, n, 2n\}$ so that the values of κ are $\kappa = 2, 1$ and $1/2$, respectively. For each combination of sample sizes (n, m) and the functional form of $g(z)$, we generate 1,000 Monte Carlo samples. The following three estimators of $\theta = (\beta_0, \beta_1, \gamma)'$ are examined: (i) the infeasible OLS (denoted by OLS* in Table 1) estimator using the unobservable complete sample $\mathcal{S}^*$; (ii) the MSOLS estimator using the matched sample $\mathcal{S}$; and (iii) the MSII-L estimator using the matched sample $\mathcal{S}$. For each estimator, averages, standard deviations (in parentheses) and root-mean squared errors (RMSEs) (in brackets) over 1,000 replications are reported.

TABLE 1 ABOUT HERE

4.2 Simulation Results

Table 1 reports simulation results for the two slope estimates to save space. Since there is only one matching variable, each of the three estimators has a $\sqrt{n}$-rate of convergence. Moreover, because of conditional homoskedasticity of the error term u, OLS* is the Gauss-Markov estimator, regardless of the specification of $g(\cdot)$ and the sample sizes. Each panel illustrates that OLS* is unbiased and yields small standard deviations. As expected, the standard deviations tend to be smaller as n increases.

However, OLS* is an infeasible, oracle estimator. Instead, we should focus on the realistic comparison between MSOLS and MSII-L, and use OLS* as the benchmark to measure the efficiency loss when all variables cannot be taken from a single data source. Table 1 illustrates that MSOLS is inconsistent and that the MSOLS estimate of β_1 is biased toward zero, as predicted. It can be also seen that the bias is non-vanishing; each panel in fact indicates that simulation averages of the MSOLS estimates do not vary across sample sizes. However, the magnitude of the bias depends on the specification of $g(\cdot)$. It can be also seen that their standard deviations shrink with n, as Theorem 1 suggests.

Now we turn to MSII-L. A first glance reveals that the bias correction works very well in that simulation averages of the MSII-L estimates are close to the truth. After a closer look, we can see that the performance of our bias correction depends on the specification of $g(\cdot)$. The estimates are often slightly biased when $n = 500$; the tendency is pronounced for Models B, C and E, which are either monotone or somewhat close to a linear function. For these models, standard deviations also tend to be large. In contrast, Models A, D and F are highly nonlinear, and for these cases biases and standard deviations are quite small even when $n = 500$. However, each panel indicates that biases vanish and standard deviations also shrink as n increases, which leads to a decrease in RMSEs and thus confirms consistency of MSII-L. It can be also seen that standard deviations for each n tend to be smaller as κ decreases, as Theorem 1 suggests.

Comparing MSII-L with OLS*, we have the following two findings. First, unlike OLS*, MSII-L is not unbiased. However, it is nearly unbiased for large sample sizes. Second, standard deviations of the latter are always greater than those of the former.

This relative efficiency loss can be thought of as the price to pay for identifying and estimating the regression using two samples jointly. The magnitude of the efficiency loss depends on the degree of nonlinearity in $g(\cdot)$. In particular, when it is close to linear (e.g., Models B, C and E), the standard deviations of MSII-L are four to six times as large as those of OLS*. In contrast, for highly nonlinear $g(\cdot)$ (e.g., Models A, D and F), the standard deviations are confined within three times those of OLS*.

To sum up, the simulation results indicate that the bias correction in MSII-L works reasonably well. While the size of the estimation error is subject to the functional form of $g(\cdot)$, it is likely to be fairly small for large samples even if the nonlinearities in $g(\cdot)$ are not very pronounced.

5 Conclusion

Regression estimation using samples constructed via the NNM from two sources is not uncommon in applied economics. This paper has demonstrated that the OLS estimation using matched samples is inconsistent and thus an appropriate bias correction is required. It has been also shown that the convergence rate to the probability limit of OLS depends on the number of matching variables. Two versions of bias-corrected estimators have been proposed, and each can be interpreted as a variant of II estimators. The MSII-L estimator attains the parametric convergence rate for the cases with only one matching variable, whereas the parametric convergence of the MSII-FD estimator can be achieved for the cases with no more than three matching variables.

Several research extensions would be fruitful. First, we may adopt the propensity score matching for further dimension reduction in matching variables. In a closely related paper, Abadie and Imbens (2012b) deliver asymptotic properties of the matching estimators for average treatment effects using an estimated propensity score as a plug-in. It is worth pursuing a similar idea for matched-sample regression estimation. Second, combining our matched-sample estimation theory with IV/GMM estimation would be also of interest in the presence of endogeneity in regressors. This is particularly relevant to empirical studies using earnings data, which are thought to include measurement errors and imputation biases. Third, the estimation theory may be extended to kernel estimation of varying coefficient models using matched samples. It is not difficult to see that kernel estimators of the varying coefficients are also inconsistent, and appropriate bias-correction methods similar to those proposed in this paper are worth investigating.

A Appendix: Technical Proofs

A.1 Useful Lemmas

Before proceeding, we provide two lemmas about the error bounds from NNM, which are repeatedly applied in the technical proofs below. The first lemma, taken from Abadie and Imbens (2006), considers NNM between two samples, whereas the second lemma, taken from Yatchew (1997), refers to NNM within the same sample.

First, we provide the formal definition of the matching discrepancy in Abadie and Imbens (2006). Let $z \in \mathbb{Z}$ be a fixed value of the matching variable Z, where, in practice, z is one of $\{Z_i\}_{i=1}^n$ in $\mathcal{S}_1$. Then, the closest matching discrepancy $U = U(z)$ is defined as $U := Z_{j(z)} - z$ if $Z_{j(z)}$ is the closest match to z among all $\{Z_j\}_{j=1}^m$ in $\mathcal{S}_2$. The following lemma states uniform moment bounds of the matching discrepancy with the number of closest neighbors M set equal to 1 in Lemma 2 of Abadie and Imbens (2006).

Lemma A1. (Abadie and Imbens, 2006, Lemma 2) *Under Assumptions 1-2, all the moments of $n^{1/d_3}\|U\|$ are uniformly bounded in n and $z \in \mathbb{Z}$.*

Second, for NNM within the same sample, the following lemma applies. In our context N is either n or m. Also observe that the ordering rule reduces to

$Z_1 \leq \cdots \leq Z_N$ when $d_3 = 1$.

Lemma A2. (Yatchew, 1997, Lemma) *Suppose that Assumptions 1-2 hold and that the support* $\mathbb{Z}$ *is the unit cube in* $\mathbb{R}^{d_3}$ *without loss of generality. Let* N *be the sample size. Select* δ *positive and arbitrarily close to* 0. *Cover the unit cube with sub-cubes of volume* $1/N^{1-\delta}$ *each with sizes* $1/N^{(1-\delta)/d_3}$. *Within each sub-cube construct a path using the nearest neighbor algorithm. Following this, knit the paths together by joining endpoints in contiguous sub-cubes to obtain the reordered sample* $\{Z_k\}_{k=1}^N$. *Then for any* $\delta > 0$, $(1/N)\sum_{k=2}^N \|\Delta Z_k\|^2 := (1/N)\sum_{k=2}^N \|Z_k - Z_{k-1}\|^2 = O_p\left(N^{-2(1-\delta)/d_3}\right)$.

A.2 Proof of Theorem 1

It is easy to see from (2) that $\hat{R}_W := \hat{Q}_W\theta + B_{R_W1} + B_{R_W2} + E_{R_W}$, where $B_{R_W1} = E\left(W_{i,j(i)}\epsilon_{i,j(i)}\right)$, $B_{R_W2} = (1/n)\sum_{i=1}^n W_{i,j(i)}\lambda_{i,j(i)}$, and $E_{R_W} = (1/n)\sum_{i=1}^n \left\{W_{i,j(i)}\epsilon_{i,j(i)} - E\left(W_{i,j(i)}\epsilon_{i,j(i)}\right)\right\}$. It follows that $\hat{\theta}_{OLS} := \theta + B_{OLS1} + B_{OLS2} + E_{OLS} + O_p\left(n^{-1/2}\right)$, where $B_{OLS1} = \hat{Q}_W^{-1}B_{R_W1}$, $B_{OLS2} = \hat{Q}_W^{-1}B_{R_W2}$ and $E_{OLS} = \hat{Q}_W^{-1}E_{R_W}$ correspond to the first-order (or leading) bias, the second-order bias due to the matching discrepancy and the weighted average of errors, respectively.

We begin with evaluating B_{OLS1}. First note that $E\left(X_{1i}\eta_{2_i}'\right) = E\left\{g_1(Z)\eta_2'\right\} + E\left(\eta_1\eta_2'\right) = 0_{d_1\times d_2}$, $E\left(X_{2j(i)}\eta_{2_{j(i)}}'\right) = \Sigma_2$, and ith and $j(i)$th observations are independent. Then, $B_{R_W1} = \left[\begin{array}{ccc} 0_{1\times d_1} & (-\Sigma_2\beta_2)' & 0_{1\times d_3}\end{array}\right]' = -\operatorname{diag}\left\{0_{d_1\times d_1}, \Sigma_2, 0_{d_3\times d_3}\right\}\theta := -\Sigma\theta$. Because $\hat{Q}_W = Q_W + O_p\left(n^{-1/2}\right)$, we obtain $B_{OLS1} = -Q_W^{-1}\Sigma\theta + O_p\left(n^{-1/2}\right)$. Next, Lemma A1 implies that $\max_{1\leq i\leq n}\left\|Z_{j(i)} - Z_i\right\| = O_p\left(n^{-1/d_3}\right)$. Then, by Cauchy-Schwarz inequality and Lipschitz continuity of g_2, $\|B_{R_W2}\|$ is bounded by $O_p\left(n^{-1/d_3}\right)$. Hence, $B_{OLS2} = O_p\left(n^{-1/d_3}\right)$. Finally, $E_{R_W} = O_p\left(n^{-1/2}\right)$ by CLT, and thus $E_{OLS} = O_p\left(n^{-1/2}\right)$. Therefore, $\hat{\theta}_{OLS} = \theta - Q_W^{-1}\Sigma\theta + O_p\left(n^{-1/d_3}\right) + O_p\left(n^{-1/2}\right) = Q_W^{-1}P_W\theta + O_p\left(n^{-\min\{1/2,1/d_3\}}\right)$ by denoting $P_W := Q_W - \Sigma$. ■

A.3 Proof of Theorem 2

Consistency of $\hat{\theta}_{II-L}$ can be established in line with the proof of Theorem 1. To derive the asymptotic distribution of $\sqrt{n}\left(\hat{\theta}_{II-L}-\theta\right)$, we first obtain

$$\hat{R}_W = \left(\hat{Q}_W - \Sigma\right)\theta + B_{R_W 2} + E_{R_W} = \hat{P}_W\theta + \left(\hat{\Sigma} - \Sigma\right)\theta + B_{R_W 2} + E_{R_W} \tag{A1}$$

by the proof of Theorem 1. Substituting this into $\sqrt{n}\left(\hat{\theta}_{II-L}-\theta\right)$ yields

$$\sqrt{n}\left(\hat{\theta}_{II-L}-\theta\right) = \hat{P}_W^{-1}\left\{\sqrt{n}\left(\hat{\Sigma}-\Sigma\right)\theta + \sqrt{n}B_{R_W 2} + \sqrt{n}E_{R_W}\right\}. \tag{A2}$$

When $d_3 = 1$, $B_{R_W 2} = O_p\left(n^{-1}\right)$ and thus $\sqrt{n}B_{R_W 2} = O_p\left(n^{-1/2}\right) = o_p(1)$. Using $\hat{P}_W^{-1} = P_W^{-1} + o_p(1)$, $\sqrt{n}\left(\hat{\Sigma}-\Sigma\right)\theta = \sqrt{n}\left(\hat{\Sigma}_2-\Sigma_2\right)\beta_2$, and $n/m = \kappa + o(1)$, we finally have

$$\sqrt{n}\left(\hat{\theta}_{II-L}-\theta\right) = P_W^{-1}\left\{\sqrt{n}E_{R_W} + \sqrt{\kappa}\sqrt{m}\left(\hat{\Sigma}_2-\Sigma_2\right)\beta_2\right\} + o_p(1),$$

where, by CLT, each of $\sqrt{n}E_{R_W}$ and $\sqrt{m}\left(\hat{\Sigma}_2-\Sigma_2\right)\beta_2$ is asymptotically normal. Therefore, $\sqrt{n}\left(\hat{\theta}_{II-L}-\theta\right) \xrightarrow{d} N\left(0, P_W^{-1}\Omega_W P_W^{-1}\right)$, where Ω_W is some $d\times d$ long-run variance matrix implied by two summands $W_{i,j(i)}\epsilon_{i,j(i)} - E\left(W_{i,j(i)}\epsilon_{i,j(i)}\right) = W_{i,j(i)}\epsilon_{i,j(i)} + \Sigma\theta$ and $\left\{\left(\Delta X_{2j}\Delta X_{2j}'/2\right)-\Sigma_2\right\}\beta_2$ in E_{R_W} and $\left(\hat{\Sigma}_2-\Sigma_2\right)\beta_2$, respectively.

The remaining task is to provide the analytical expression of Ω_W. Observe that Ω_W may be rewritten as

$$\Omega_W := \Omega_{W,11} + \sqrt{\kappa}\left(\Omega_{W,12}+\Omega_{W,12}'\right) + \kappa\Omega_{W,22},$$

where $\Omega_{W,11}$ and $\Omega_{W,22}$ are the long-run variance matrices of $W_{i,j(i)}\epsilon_{i,j(i)} + \Sigma\theta$ and $\left\{\left(\Delta X_{2j}\Delta X_{2j}'/2\right)-\Sigma_2\right\}\beta_2$, respectively, and $\Omega_{W,12}$ is their long-run covariance. Now, let $\phi_{i,j(i)} := W_{i,j(i)}\epsilon_{i,j(i)} + \Sigma\theta$. Clearly, this has no serial dependence, and thus

$$\Omega_{W,11} = Var\left(\phi_{i,j(i)}\right) = E\left(\phi_{i,j(i)}\phi_{i,j(i)}'\right).$$

Next, similarly to Lemma A3, $\hat{\Sigma}_2 = (m-1)^{-1}\sum_{j=2}^{m}\left(\Delta\eta_{2j}\Delta\eta_{2j}'/2\right) + o_p\left(m^{-1/2}\right)$ as long as $d_3 \le 3$, where $E\left(\Delta\eta_{2j}\Delta\eta_{2j}'/2\right) = \Sigma_2$. Because $\psi_j := \left\{\left(\Delta\eta_{2j}\Delta\eta_{2j}'/2\right)-\Sigma_2\right\}\beta_2$

is one-dependent, its long-run variance is $E\left(\psi_j\psi_{j-1}'\right)+E\left(\psi_j\psi_j'\right)+E\left(\psi_j\psi_{j+1}'\right)$ and thus

$$\Omega_{W,22}=\text{diag}\left\{0_{d_1\times d_1},E\left(\psi_j\psi_{j-1}'\right)+E\left(\psi_j\psi_j'\right)+E\left(\psi_j\psi_{j+1}'\right),0_{d_3\times d_3}\right\}.$$

Lastly, for each i, $\eta_{2j(i)}$ in $\phi_{i,j(i)}$ is correlated with $\psi_{j(i)}$ and $\psi_{j(i+1)}$. Now we decompose $\psi_{j(i)}$ as follows

$$\begin{aligned}\psi_{j(i)} &= \left\{\left(\eta_{2j(i)}\eta_{2j(i)}'-\Sigma_2\right)/2+\left(\eta_{2j(i-1)}\eta_{2j(i-1)}'-\Sigma_2\right)/2\right.\\ &\quad\left.-\eta_{2j(i)}\eta_{2j(i-1)}'-\eta_{2j(i-1)}\eta_{2j(i)}'\right\}\beta_2\end{aligned}$$

Because ith and $j\left(i\right)$th observations are independent, only the expectation between $\phi_{i,j(i)}$ and the first term is non-zero. Defining $\omega_{j(i)}:=\left(\eta_{2j(i)}\eta_{2j(i)}'-\Sigma_2\right)\beta_2$, we have $E\left(\phi_{i,j(i)}\psi_{j(i)}'\right)=E\left(\phi_{i,j(i)}\omega_{j(i)}'\right)/2$. Similarly, $E\left(\phi_{i,j(i)}\psi_{j(i+1)}'\right)=E\left(\phi_{i,j(i)}\omega_{j(i)}'\right)/2$ also holds. Therefore,

$$\begin{aligned}\Omega_{W,12}&=\left[\begin{array}{ccc}0_{d\times d_1} & E\left(\phi_{i,j(i)}\psi_{j(i)}'\right)+E\left(\phi_{i,j(i)}\psi_{j(i+1)}'\right) & 0_{d\times d_3}\end{array}\right]\\ &=\left[\begin{array}{ccc}0_{d\times d_1} & E\left(\phi_{i,j(i)}\omega_{j(i)}'\right) & 0_{d\times d_3}\end{array}\right],\end{aligned}$$

which completes the proof. ■

A.4 Proof of Theorem 3

The proof largely follows from the one of Proposition 1 in Yatchew (1997). Before proceeding, the proof requires the following lemma.

Lemma A3. *Let* $g_{i,j(i)}:=\left(g_1\left(Z_i\right)',g_{2\backslash 0}\left(Z_{j(i)}\right)'\right)'$ *and* $\eta_{i,j(i)}:=\left(\eta_{1i}',\eta_{2\backslash 0j(i)}'\right)'$. *Then, under Assumptions 1-3,*

$$\frac{1}{n-1}\sum_{i=2}^{n}\Delta g_{i,j(i)}\Delta\mu_{i,j(i)}=O_p\left(n^{-2(1-\delta)/d_3}\right),\tag{A3}$$

$$\frac{1}{n-1}\sum_{i=2}^{n}\Delta\eta_{i,j(i)}\Delta\mu_{i,j(i)}=o_p\left(n^{-1/2}\right),\ \textit{and}\tag{A4}$$

$$\frac{1}{n-1}\sum_{i=2}^{n}\Delta g_{i,j(i)}\Delta\epsilon_{i,j(i)}=o_p\left(n^{-1/2}\right),\tag{A5}$$

where $\delta\left(>0\right)$ *is a constant arbitrarily close to* 0.

A.4.1 Proof of Lemma A3

By Lipschitz continuity of $g(\cdot)$,

$$\begin{aligned}\|\Delta g_1(Z_i)\| &\leq c\|\Delta Z_i\|,\\ \left\|\Delta g_{2\backslash 0}\left(Z_{j(i)}\right)\right\| &\leq c\left(\|\Delta Z_i\| + \left\|Z_{j(i)} - Z_i\right\| + \left\|Z_{j(i-1)} - Z_{i-1}\right\|\right), \text{ and}\\ \left\|\Delta \mu_{i,j(i)}\right\| &\leq c\left(\|\Delta Z_i\| + \left\|Z_{j(i)} - Z_i\right\| + \left\|Z_{j(i-1)} - Z_{i-1}\right\|\right).\end{aligned}$$

Then, (A3) can be established by applying Cauchy-Schwarz inequality and Lemmata A1-A2.

Furthermore, to show (A4), notice that

$$E\left(\Delta\eta_{i,j(i)}\Delta\mu_{i,j(i)}\right) = E\left\{E\left(\Delta\eta_{i,j(i)}\middle| Z_i, Z_{i-1}, Z_{j(i)}, Z_{j(i-1)}\right)\Delta\mu_{i,j(i)}\right\} = 0.$$

To evaluate the order of magnitude in $Var\left\{(n-1)^{-1}\sum_{i=2}^n \Delta\eta_{i,j(i)}\Delta\mu_{i,j(i)}\right\}$, consider that among all summands of $(n-1)^{-2}\sum_{i=2}^n\sum_{k=2}^n \Delta\eta_{i,j(i)}\Delta\eta'_{k,j(k)}\Delta\mu_{i,j(i)}\Delta\mu_{k,j(k)}$, only $\Delta\eta_{i,j(i)}\Delta\eta'_{i,j(i)}\left(\Delta\mu_{i,j(i)}\right)^2$, $\Delta\eta_{i,j(i)}\Delta\eta'_{i-1,j(i-1)}\Delta\mu_{i,j(i)}\Delta\mu_{i-1,j(i-1)}$ and $\Delta\eta_{i,j(i)}\Delta\eta'_{i+1,j(i+1)}\Delta\mu_{i,j(i)}\Delta\mu_{i+1,j(i+1)}$ have non-zero expectations. Because, for each of $k = i, i \pm 1$,

$$\frac{1}{(n-1)^2}\left\|\sum_{i=2}^n \Delta\eta_{i,j(i)}\Delta\eta'_{k,j(k)}\Delta\mu_{i,j(i)}\Delta\mu_{k,j(k)}\right\| \leq O_p\left(n^{-\{1+2(1-\delta)/d_3\}}\right),$$

$Var\left\{(n-1)^{-1}\sum_{i=2}^n \Delta\eta_{i,j(i)}\Delta\mu_{i,j(i)}\right\}$ is bounded by $O_p\left(n^{-\{1+2(1-\delta)/d_3\}}\right) = o_p\left(n^{-1}\right)$ and thus (A4) follows. Likewise, (A5) can be also demonstrated. ■

A.4.2 Proof of Theorem 3

Using $\Delta X_{i,j(i)} = \Delta g_{i,j(i)} + \Delta\eta_{i,j(i)}$, $\Delta\epsilon_{i,j(i)} = \Delta u_i + \left(\Delta\eta_{2\backslash 0i} - \Delta\eta_{2\backslash 0j(i)}\right)'\beta_{2\backslash 0}$, and Lemma A3, we may write $\hat{R}_{\Delta X} := \hat{Q}_{\Delta X}\beta + B_{R_{\Delta X}1} + B_{R_{\Delta X}2} + E_{R_{\Delta X}}$, where

$$\begin{aligned}
B_{R_{\Delta X}1} &= E\left(\Delta\eta_{i,j(i)}\Delta\epsilon_{i,j(i)}\right) = \begin{bmatrix} 0_{d_1\times 1} \\ -2\Sigma_{2\backslash 0}\beta_{2\backslash 0} \end{bmatrix} = -2\,\mathrm{diag}\left\{0_{d_1\times d_1}, \Sigma_{2\backslash 0}\right\}\beta := -2\Sigma_{FD}\beta, \\
B_{R_{\Delta X}2} &= \frac{1}{n-1}\left(\sum_{i=2}^{n}\Delta g_{i,j(i)}\Delta\mu_{i,j(i)} + \sum_{i=2}^{n}\Delta\eta_{i,j(i)}\Delta\mu_{i,j(i)} + \sum_{i=2}^{n}\Delta g_{i,j(i)}\Delta\epsilon_{i,j(i)}\right) \\
&= O_p\left(n^{-2(1-\delta)/d_3}\right) + o_p\left(n^{-1/2}\right), \text{ and} \\
E_{R_{\Delta X}} &= \frac{1}{n-1}\sum_{i=2}^{n}\left\{\Delta\eta_{i,j(i)}\Delta\epsilon_{i,j(i)} - E\left(\Delta\eta_{i,j(i)}\Delta\epsilon_{i,j(i)}\right)\right\} = O_p\left(n^{-1/2}\right).
\end{aligned}$$

Then, the result immediately follows from $\hat{Q}_{\Delta X} = Q_{\Delta X} + O_p\left(n^{-1/2}\right)$. ■

A.5 Proof of Proposition 2

Consistency of each estimator is obvious. For asymptotic normality, we only demonstrate that

$$\sqrt{n}\left(\hat{\theta}_{FM-L} - \theta\right) = \sqrt{n}\left(\hat{\theta}_{II-L} - \theta\right) + o_p(1) \tag{A6}$$

to save space. By (A1) and (A2),

$$\begin{aligned}
\hat{R}_W + \hat{\Sigma}\hat{\theta}_{II-L} &= \hat{Q}_W\theta + \Sigma\left(\hat{\theta}_{II-L} - \theta\right) + \left\{\left(\hat{\Sigma} - \Sigma\right)\theta + B_{R_W2} + E_{R_W}\right\} + o_p\left(n^{-1/2}\right) \\
&= \hat{Q}_W\theta + \left(\hat{P}_W + \Sigma\right)\left(\hat{\theta}_{II-L} - \theta\right) + o_p\left(n^{-1/2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_{FM-L} - \theta\right) &= \hat{Q}_W^{-1}\left\{\hat{Q}_W - \left(\hat{\Sigma} - \Sigma\right)\right\}\sqrt{n}\left(\hat{\theta}_{II-L} - \theta\right) + o_p(1) \\
&= \left\{I_d + O_p\left(m^{-1/2}\right)\right\}\sqrt{n}\left(\hat{\theta}_{II-L} - \theta\right) + o_p(1),
\end{aligned}$$

and thus (A6) indeed holds. ■

References

[1] Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235 - 267.

[2] Abadie, A., and G. W. Imbens (2012a): "A Martingale Representation for Matching Estimators," *Journal of the American Statistical Association*, 107, 833 - 843.

[3] Abadie, A., and G. W. Imbens (2012b): "Matching on the Estimated Propensity Score," Working Paper, John F. Kennedy School of Government, Harvard University.

[4] Angrist, J. D., and A. B. Krueger (1992): "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 87, 328 - 336.

[5] Angrist, J. D., and A. B. Krueger (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business & Economic Statistics*, 13, 225 - 235.

[6] Arellano, M., and C. Meghir (1992): "Female Labour Supply and on the Job Search: An Empirical Model Estimated Using Complementary Data Sets," *Review of Economic Studies*, 59, 537 - 559.

[7] Björklund, J., and M. Jäntti (1997): "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review*, 87, 1009 - 1018.

[8] Bollinger, C. R., and B. T. Hirsch (2006): "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics*, 24, 483 - 519.

[9] Borjas, G. J. (2004): "Food Insecurity and Public Assistance," *Journal of Public Economics*, 88, 1421 - 1443.

[10] Busso, M., J. DiNardo, and J. McCrary (2014): "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators," *Review of Economics and Statistics*, in press.

[11] Carrasco, M., and J.-P. Florens (2002): "Simulation-Based Method of Moments and Efficiency," *Journal of Business & Economic Statistics*, 20, 482 - 492.

[12] Currie, J., and A. Yelowitz (2000): "Are Public Housing Projects Good for Kids?" *Journal of Public Economics*, 75, 99 - 124.

[13] Dee, T. S., and W. N. Evans (2003): "Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates," *Journal of Labor Economics*, 21, 178 - 209.

[14] Fujii, T. (2008): "Two-Sample Estimation of Poverty Rates for Disabled People: An Application to Tanzania," Singapore Management University Economics & Statistics Working Paper No.02-2008.

[15] Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85 - S118.

[16] Hellerstein, J. K., and G. W. Imbens (1999): "Imposing Moment Restrictions from Auxiliary Data by Weighting," *Review of Economics and Statistics*, 81, 1 - 14.

[17] Hirsch, B. T., and E. J. Schumacher (2004): "Match Bias in Wage Gap Estimates due to Earnings Imputation," *Journal of Labor Economics*, 22, 689 - 722.

[18] Horowitz, J. L., and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599 - 631.

[19] Imbens, G. W., and T. Lancaster (1994): "Combining Micro and Marco Data in Microeconometric Models," *Review of Economic Studies*, 61, 655 - 680.

[20] Inoue, A., and G. Solon (2010): "Two-Sample Instrumental Variables Estimators," *Review of Economics and Statistics*, 92, 557 - 561.

[21] Little, R. J. A., and D. B. Rubin (2002): *Statistical Analysis with Missing Data*, Second Edition. New York: John Wiley & Sons.

[22] Lusardi, A. (1996): "Permanent Income, Current Income, and Consumption: Evidence from Two Panel Data Sets," *Journal of Business & Economic Statistics*, 14, 81 - 90.

[23] Murtazashvili, I., D. Liu, and A. Prokhorov (2013): "Two-Sample Nonparametric Estimation of Intergenerational Income Mobility," Business Analytics Working Paper No: 07/2013, University of Sydney.

[24] Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221 - 247.

[25] Phillips, P. C. B., and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99 - 125.

[26] Prokhorov, A., and P. Schmidt (2009): "GMM Redundancy Results for General Missing Data Problems," *Journal of Econometrics*, 151, 47 - 55.

[27] Rice, J. (1984): "Bandwidth Choice for Nonparametric Regression," *Annals of Statistics*, 12, 1215 - 1230.

[28] Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931 - 954.

[29] Smith, A. A., Jr. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63 - S84.

[30] Smith, J. A., and P. E. Todd (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators," *Journal of Econometrics*, 125, 305 - 353.

[31] Yatchew, A. (1997): "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, 57, 135 - 143.

Table 1: Monte Carlo Results

Model A: $g(z) = z^2$

n	Estimator	$m = n/2$ $(\kappa = 2)$		$m = n$ $(\kappa = 1)$		$m = 2n$ $(\kappa = 1/2)$	
		β_1	γ	β_1	γ	β_1	γ
500	OLS*	0.9986	0.9993	0.9978	1.0003	0.9998	1.0007
		(0.0292)	(0.0378)	(0.0300)	(0.0400)	(0.0295)	(0.0391)
		[0.0292]	[0.0378]	[0.0301]	[0.0400]	[0.0295]	[0.0391]
	MSOLS	0.5930	1.0016	0.5867	1.0002	0.5878	1.0000
		(0.0513)	(0.0741)	(0.0498)	(0.0708)	(0.0483)	(0.0654)
		[0.4102]	[0.0741]	[0.4163]	[0.0708]	[0.4150]	[0.0654]
	MSII-L	1.0338	1.0028	1.0152	0.9993	1.0114	0.9976
		(0.1455)	(0.0971)	(0.1184)	(0.0838)	(0.1071)	(0.0752)
		[0.1493]	[0.0972]	[0.1193]	[0.0838]	[0.1077]	[0.0752]
1000	OLS*	0.9998	1.0011	1.0000	1.0001	0.9999	1.0009
		(0.0202)	(0.0282)	(0.0211)	(0.0275)	(0.0207)	(0.0272)
		[0.0202]	[0.0283]	[0.0211]	[0.0276]	[0.0207]	[0.0272]
	MSOLS	0.5884	1.0006	0.5885	1.0001	0.5882	1.0018
		(0.0360)	(0.0536)	(0.0342)	(0.0487)	(0.0338)	(0.0479)
		[0.4131]	[0.0536]	[0.4130]	[0.0487]	[0.4131]	[0.0479]
	MSII-L	1.0090	0.9989	1.0139	1.0015	1.0099	1.0016
		(0.0927)	(0.0681)	(0.0794)	(0.0576)	(0.0718)	(0.0542)
		[0.0931]	[0.0681]	[0.0806]	[0.0576]	[0.0725]	[0.0543]
2000	OLS*	0.9990	0.9991	0.9995	0.9998	0.9995	1.0004
		(0.0148)	(0.0186)	(0.0143)	(0.0193)	(0.0144)	(0.0192)
		[0.0148]	[0.0186]	[0.0143]	[0.0193]	[0.0144]	[0.0192]
	MSOLS	0.5877	0.9997	0.5869	0.9985	0.5866	1.0005
		(0.0262)	(0.0379)	(0.0244)	(0.0349)	(0.0236)	(0.0321)
		[0.4131]	[0.0379]	[0.4138]	[0.0349]	[0.4141]	[0.0321]
	MSII-L	1.0062	0.9996	1.0019	0.9984	1.0040	1.0006
		(0.0650)	(0.0484)	(0.0551)	(0.0413)	(0.0528)	(0.0358)
		[0.0653]	[0.0484]	[0.0551]	[0.0413]	[0.0529]	[0.0358]

Table 1: *Continued*

Model B: $g(z) = z^3$

n	Estimator	$m = n/2$ $(\kappa = 2)$		$m = n$ $(\kappa = 1)$		$m = 2n$ $(\kappa = 1/2)$	
		β_1	γ	β_1	γ	β_1	γ
500	OLS*	1.0002	0.9988	0.9991	1.0026	1.0000	1.0008
		(0.0289)	(0.0784)	(0.0292)	(0.0815)	(0.0273)	(0.0764)
		[0.0289]	[0.0784]	[0.0292]	[0.0816]	[0.0273]	[0.0764]
	MSOLS	0.6002	1.9593	0.5957	1.9695	0.5949	1.9705
		(0.0531)	(0.1435)	(0.0482)	(0.1317)	(0.0463)	(0.1255)
		[0.4033]	[0.9700]	[0.4072]	[0.9784]	[0.4078]	[0.9786]
	MSII-L	1.0369	0.9153	1.0209	0.9499	1.0103	0.9728
		(0.1438)	(0.3586)	(0.1106)	(0.2787)	(0.0988)	(0.2485)
		[0.1485]	[0.3684]	[0.1125]	[0.2832]	[0.0993]	[0.2500]
1000	OLS*	0.9996	1.0020	0.9999	1.0004	1.0009	0.9987
		(0.0203)	(0.0575)	(0.0205)	(0.0574)	(0.0205)	(0.0565)
		[0.0203]	[0.0575]	[0.0205]	[0.0574]	[0.0206]	[0.0566]
	MSOLS	0.5952	1.9720	0.5953	1.9708	0.5954	1.9712
		(0.0371)	(0.1016)	(0.0331)	(0.0886)	(0.0342)	(0.0910)
		[0.4065]	[0.9773]	[0.4061]	[0.9749]	[0.4061]	[0.9755]
	MSII-L	1.0108	0.9733	1.0119	0.9731	1.0083	0.9816
		(0.0901)	(0.2275)	(0.0733)	(0.1838)	(0.0711)	(0.1774)
		[0.0907]	[0.2291]	[0.0742]	[0.1857]	[0.0716]	[0.1783]
2000	OLS*	0.9997	0.9999	0.9998	1.0002	0.9996	1.0014
		(0.0144)	(0.0388)	(0.0146)	(0.0404)	(0.0138)	(0.0384)
		[0.0144]	[0.0388]	[0.0146]	[0.0404]	[0.0138]	[0.0384]
	MSOLS	0.5945	1.9724	0.5948	1.9707	0.5933	1.9767
		(0.0265)	(0.0720)	(0.0240)	(0.0670)	(0.0231)	(0.0635)
		[0.4064]	[0.9750]	[0.4059]	[0.9730]	[0.4074]	[0.9788]
	MSII-L	1.0059	0.9854	1.0053	0.9856	1.0019	0.9960
		(0.0650)	(0.1615)	(0.0548)	(0.1403)	(0.0507)	(0.1276)
		[0.0653]	[0.1622]	[0.0551]	[0.1411]	[0.0507]	[0.1276]

Table 1: *Continued*

Model C: $g(z) = \exp(z)$

		$m = n/2$ $(\kappa = 2)$		$m = n$ $(\kappa = 1)$		$m = 2n$ $(\kappa = 1/2)$	
n	Estimator	β_1	γ	β_1	γ	β_1	γ
500	OLS*	0.9986	1.0012	0.9974	1.0042	0.9998	1.0010
		(0.0349)	(0.0635)	(0.0358)	(0.0667)	(0.0351)	(0.0654)
		[0.0349]	[0.0635]	[0.0359]	[0.0668]	[0.0351]	[0.0654]
	MSOLS	0.4123	1.8591	0.4052	1.8690	0.4076	1.8660
		(0.0583)	(0.1044)	(0.0584)	(0.1023)	(0.0569)	(0.0993)
		[0.5906]	[0.8655]	[0.5976]	[0.8750]	[0.5951]	[0.8716]
	MSII-L	1.1123	0.8402	1.0505	0.9258	1.0328	0.9500
		(0.4245)	(0.6211)	(0.2455)	(0.3677)	(0.2008)	(0.3029)
		[0.4391]	[0.6413]	[0.2507]	[0.3751]	[0.2034]	[0.3070]
1000	OLS*	0.9996	1.0017	0.9999	1.0003	1.0002	1.0006
		(0.0241)	(0.0452)	(0.0253)	(0.0467)	(0.0250)	(0.0458)
		[0.0241]	[0.0452]	[0.0253]	[0.0467]	[0.0250]	[0.0458]
	MSOLS	0.4072	1.8673	0.4068	1.8662	0.4076	1.8669
		(0.0412)	(0.0743)	(0.0400)	(0.0685)	(0.0398)	(0.0712)
		[0.5942]	[0.8705]	[0.5945]	[0.8689]	[0.5937]	[0.8699]
	MSII-L	1.0295	0.9557	1.0320	0.9548	1.0236	0.9671
		(0.1799)	(0.2743)	(0.1463)	(0.2205)	(0.1312)	(0.1977)
		[0.1823]	[0.2778]	[0.1497]	[0.2251]	[0.1333]	[0.2004]
2000	OLS*	0.9988	1.0008	0.9994	1.0006	0.9994	1.0012
		(0.0176)	(0.0320)	(0.0171)	(0.0320)	(0.0172)	(0.0318)
		[0.0177]	[0.0320]	[0.0171]	[0.0320]	[0.0172]	[0.0318]
	MSOLS	0.4067	1.8666	0.4067	1.8656	0.4055	1.8693
		(0.0300)	(0.0536)	(0.0282)	(0.0505)	(0.0278)	(0.0492)
		[0.5941]	[0.8682]	[0.5940]	[0.8670]	[0.5952]	[0.8706]
	MSII-L	1.0164	0.9757	1.0084	0.9861	1.0094	0.9869
		(0.1211)	(0.1832)	(0.0993)	(0.1521)	(0.0960)	(0.1459)
		[0.1222]	[0.1848]	[0.0996]	[0.1528]	[0.0964]	[0.1465]

Table 1: *Continued*

Model D: $g(z) = z + 2\sin(\pi z)$

n	Estimator	$m = n/2$ $(\kappa = 2)$ β_1	$m = n/2$ $(\kappa = 2)$ γ	$m = n$ $(\kappa = 1)$ β_1	$m = n$ $(\kappa = 1)$ γ	$m = 2n$ $(\kappa = 1/2)$ β_1	$m = 2n$ $(\kappa = 1/2)$ γ
500	OLS*	1.0004	0.9991	0.9995	1.0006	1.0001	1.0007
		(0.0278)	(0.0401)	(0.0276)	(0.0434)	(0.0271)	(0.0414)
		[0.0278]	[0.0401]	[0.0276]	[0.0434]	[0.0271]	[0.0414]
	MSOLS	0.6381	1.1904	0.6320	1.1919	0.6317	1.1917
		(0.0495)	(0.0787)	(0.0483)	(0.0739)	(0.0461)	(0.0685)
		[0.3653]	[0.2060]	[0.3712]	[0.2056]	[0.3712]	[0.2036]
	MSII-L	1.0333	0.9851	1.0158	0.9911	1.0129	0.9907
		(0.1196)	(0.1171)	(0.1033)	(0.1011)	(0.0922)	(0.0896)
		[0.1242]	[0.1181]	[0.1045]	[0.1015]	[0.0931]	[0.0901]
1000	OLS*	0.9996	1.0013	0.9996	1.0004	0.9994	1.0012
		(0.0187)	(0.0296)	(0.0189)	(0.0296)	(0.0191)	(0.0288)
		[0.0188]	[0.0297]	[0.0189]	[0.0296]	[0.0191]	[0.0288]
	MSOLS	0.6303	1.1939	0.6295	1.1948	0.6305	1.1942
		(0.0353)	(0.0562)	(0.0330)	(0.0491)	(0.0312)	(0.0496)
		[0.3714]	[0.2019]	[0.3719]	[0.2009]	[0.3708]	[0.2004]
	MSII-L	1.0083	0.9946	1.0060	0.9985	1.0072	0.9975
		(0.0805)	(0.0813)	(0.0718)	(0.0666)	(0.0620)	(0.0634)
		[0.0809]	[0.0815]	[0.0721]	[0.0666]	[0.0624]	[0.0635]
2000	OLS*	0.9991	0.9996	0.9997	1.0000	1.0003	1.0002
		(0.0135)	(0.0204)	(0.0135)	(0.0208)	(0.0136)	(0.0204)
		[0.0136]	[0.0204]	[0.0135]	[0.0208]	[0.0136]	[0.0204]
	MSOLS	0.6300	1.1932	0.6301	1.1922	0.6293	1.1942
		(0.0246)	(0.0399)	(0.0236)	(0.0362)	(0.0231)	(0.0332)
		[0.3708]	[0.1972]	[0.3706]	[0.1956]	[0.3714]	[0.1970]
	MSII-L	1.0035	0.9979	1.0041	0.9963	1.0036	0.9988
		(0.0550)	(0.0563)	(0.0491)	(0.0496)	(0.0447)	(0.0425)
		[0.0551]	[0.0563]	[0.0493]	[0.0498]	[0.0448]	[0.0426]

Table 1: *Continued*

Model E: $g(z) = z + (5/\tau)\,\phi(z/\tau)\,, \tau = 0.9$

		$m = n/2$ $(\kappa = 2)$		$m = n$ $(\kappa = 1)$		$m = 2n$ $(\kappa = 1/2)$	
n	Estimator	β_1	γ	β_1	γ	β_1	γ
500	OLS*	1.0006	0.9987	0.9991	1.0013	1.0001	1.0006
		(0.0361)	(0.0523)	(0.0377)	(0.0561)	(0.0374)	(0.0538)
		[0.0361]	[0.0523]	[0.0377]	[0.0561]	[0.0374]	[0.0538]
	MSOLS	0.3539	1.6457	0.3481	1.6518	0.3474	1.6510
		(0.0615)	(0.0908)	(0.0586)	(0.0882)	(0.0560)	(0.0838)
		[0.6490]	[0.6521]	[0.6545]	[0.6577]	[0.6550]	[0.6564]
	MSII-L	1.1524	0.8469	1.0852	0.9142	1.0652	0.9321
		(0.5282)	(0.5627)	(0.3125)	(0.3283)	(0.2794)	(0.2902)
		[0.5497]	[0.5832]	[0.3239]	[0.3393]	[0.2869]	[0.2980]
1000	OLS*	0.9991	1.0020	1.0000	1.0002	0.9995	1.0014
		(0.0253)	(0.0385)	(0.0252)	(0.0381)	(0.0250)	(0.0357)
		[0.0253]	[0.0385]	[0.0252]	[0.0381]	[0.0250]	[0.0357]
	MSOLS	0.3449	1.6557	0.3458	1.6549	0.3466	1.6536
		(0.0430)	(0.0622)	(0.0399)	(0.0582)	(0.0382)	(0.0572)
		[0.6565]	[0.6586]	[0.6554]	[0.6575]	[0.6545]	[0.6561]
	MSII-L	1.0543	0.9444	1.0357	0.9659	1.0297	0.9719
		(0.2480)	(0.2583)	(0.1935)	(0.2005)	(0.1709)	(0.1771)
		[0.2539]	[0.2642]	[0.1968]	[0.2034]	[0.1734]	[0.1793]
2000	OLS*	0.9999	0.9993	1.0004	0.9994	1.0009	0.9995
		(0.0177)	(0.0261)	(0.0191)	(0.0280)	(0.0181)	(0.0261)
		[0.0177]	[0.0261]	[0.0191]	[0.0280]	[0.0182]	[0.0261]
	MSOLS	0.3464	1.6531	0.3465	1.6528	0.3456	1.6549
		(0.0309)	(0.0453)	(0.0287)	(0.0438)	(0.0276)	(0.0403)
		[0.6543]	[0.6547]	[0.6541]	[0.6543]	[0.6549]	[0.6562]
	MSII-L	1.0220	0.9778	1.0227	0.9756	1.0144	0.9863
		(0.1491)	(0.1554)	(0.1325)	(0.1418)	(0.1170)	(0.1218)
		[0.1508]	[0.1570]	[0.1344]	[0.1439]	[0.1179]	[0.1226]

Table 1: *Continued*

Model F: $g(z) = z + (5/\tau)\,\phi(z/\tau)\,, \tau = 0.3$

		$m = n/2$ ($\kappa = 2$)		$m = n$ ($\kappa = 1$)		$m = 2n$ ($\kappa = 1/2$)	
n	Estimator	β_1	γ	β_1	γ	β_1	γ
500	OLS*	1.0001	0.9992	0.9997	1.0007	0.9998	1.0008
		(0.0120)	(0.0398)	(0.0117)	(0.0419)	(0.0118)	(0.0408)
		[0.0120]	[0.0398]	[0.0117]	[0.0419]	[0.0118]	[0.0408]
	MSOLS	0.9363	1.0659	0.9363	1.0635	0.9357	1.0616
		(0.0280)	(0.0947)	(0.0235)	(0.0846)	(0.0225)	(0.0757)
		[0.0696]	[0.1154]	[0.0679]	[0.1057]	[0.0681]	[0.0975]
	MSII-L	1.0035	0.9990	1.0027	0.9969	1.0015	0.9958
		(0.0315)	(0.0995)	(0.0265)	(0.0881)	(0.0245)	(0.0782)
		[0.0317]	[0.0995]	[0.0266]	[0.0882]	[0.0246]	[0.0783]
1000	OLS*	0.9997	1.0014	0.9997	1.0005	0.9999	1.0010
		(0.0081)	(0.0295)	(0.0079)	(0.0287)	(0.0079)	(0.0278)
		[0.0081]	[0.0295]	[0.0079]	[0.0287]	[0.0079]	[0.0278]
	MSOLS	0.9365	1.0625	0.9338	1.0677	0.9355	1.0656
		(0.0191)	(0.0671)	(0.0166)	(0.0565)	(0.0154)	(0.0541)
		[0.0663]	[0.0917]	[0.0683]	[0.0882]	[0.0663]	[0.0851]
	MSII-L	1.0025	0.9964	0.9993	1.0023	1.0011	1.0005
		(0.0220)	(0.0707)	(0.0185)	(0.0590)	(0.0169)	(0.0557)
		[0.0221]	[0.0708]	[0.0185]	[0.0590]	[0.0170]	[0.0557]
2000	OLS*	1.0001	0.9990	1.0004	0.9994	1.0004	1.0000
		(0.0057)	(0.0194)	(0.0059)	(0.0203)	(0.0060)	(0.0197)
		[0.0057]	[0.0194]	[0.0059]	[0.0203]	[0.0060]	[0.0197]
	MSOLS	0.9353	1.0644	0.9358	1.0627	0.9352	1.0654
		(0.0138)	(0.0475)	(0.0115)	(0.0422)	(0.0109)	(0.0363)
		[0.0661]	[0.0800]	[0.0652]	[0.0756]	[0.0657]	[0.0748]
	MSII-L	1.0005	0.9992	1.0010	0.9974	1.0005	1.0002
		(0.0155)	(0.0497)	(0.0130)	(0.0441)	(0.0121)	(0.0374)
		[0.0155]	[0.0497]	[0.0131]	[0.0442]	[0.0121]	[0.0374]

Note: For each estimator, simulation averages of estimates, simulation standard deviations (in parentheses) and RMSEs (in brackets) are presented.