TECHNICAL WORKING PAPER SERIES

WHEN TO CONTROL FOR COVARIATES? PANEL-ASYMPTOTIC RESULTS FOR ESTIMATES OF TREATMENT EFFECTS

Joshua D. Angrist
Jinyong Hahn

Technical Working Paper 241
http://www.nber.org/papers/T0241

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1999

We are grateful to Alberto Abadie, Gary Chamberlain, Guido Imbens, and seminar participants at Penn State University and MIT econometrics luch for helpful suggestions and comments. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects

Joshua D. Angrist and Jinyong Hahn

NBER Technical Working Paper No. 241
May 1999
JEL No. C14, C23, J31

ABSTRACT

The problem of how to control for covariates is endemic in evaluation research. Covariate-matching provides an appealing control strategy, but with continuous or high-dimensional covariate vectors, exact matching may be impossible or involve small cells. Matching observations that have the same propensity score produces unbiased estimates of causal effects whenever covariate-matching does, and also has an attractive dimension-reducing property. On the other hand, conventional asymptotic arguments show that covariate-matching is (asymptotically) more efficient that propensity score-matching. This is because the usual asymptotic sequence has cell sizes growing to infinity, with no benefit from reducing the number of cells. Here, we approximate the large sample behavior of difference matching estimators using a panel-style asymptotic sequence with fixed cell sizes and the number of cells increasing to infinity. Exact calculations in simple examples and Monte Carlo evidence suggests this generates a substantially improved approximation to actual finite-sample distributions. Under this sequence, propensity-score-matching is most likely to dominate exact matching when cell sizes are small, the explanatory power of the covariates conditional on the propensity score is low, and/or the probability of treatment is close to zero or one. Finally, we introduce a random-effects type combination estimator that provides finite-sample efficiency gains over both covariate-matching and propensity-score-matching.

Joshua D. Angrist
MIT
Economics Department
50 Memorial Drive
Cambridge, MA 02138
and NBER
angrist@mit.edu

Jinyong Hahn
Univeristy of Pennsylvania
Economics Department
3718 Locust Walk
Philadelphia, PA 19104
hahn@econ.sas.upenn.edu

1 Introduction

Evaluation research typically begins with treatment-control comparisons. For example, estimates of the effect of training programs on earnings compare the earnings of those who receive training with a candidate control sample of untrained people. Because trainees are not chosen randomly, candidate control samples may not provide a very accurate picture of what would have happened to the trainees had they not been trained. This motivates attempts to reduce and perhaps even eliminate bias by controlling for covariates. Examples of training program evaluations in this spirit include Ashenfelter and Card (1985), Card and Sullivan (1988), Dehejia and Wahba (1998), and Heckman, Ichimura, and Todd (1997), all of which estimate the effects of training programs on earnings or employment after conditioning on an array of personal characteristics, including earnings and/or employment histories. Similarly, Angrist (1998) estimates the effect of voluntary service on the earnings of military applicants by conditioning on the personal characteristics used by military recruiters to select soldiers.

A problem that often arises in studies of this type is how to control for continuously distributed or high-dimensional covariates. In many training evaluations, for example, the sample sizes are small, there are many covariates, and some of the covariates, such as past earnings, are continuous. This leads to small or missing covariate-cells. A number of variations on exact covariate-matching schemes have been developed to deal with situations like this. Typically these involve approximate matching or non-parametric smoothing of some kind.[1] A practical problem with strategies of this type is that even though different estimators may have very different properties, the existing theory provides little in the way of specific guidelines as to how to choose between them.

An alternative strategy to control for covariates begins with Rosenbaum and Rubin's (1983) observation that bias from covariates can be eliminated by controlling for a scalar-valued function of the covariates, the propensity score. For a formal statement of this result, denote the covariate vector for person i by X_i treatment status by D_i, and define the conditional probability of treatment, or propensity score, by $p(X_i) \equiv \Pr[D_i = 1 | X_i]$. Let Y_{0i} denote the potential or counter-factual earnings of a trainee if he or she had not been trained, and let Y_{1i} denote potential earnings as a trainee. The assumption that motivates

[1] See, e.g., Cochran (1965), Rubin (1973, 1979), or Rosenbaum (1995, Chapter 9), for discussions of caliper and nearest-neighbor matching, and Deaton and Paxson (1998) or Heckman, Ichimura and Todd (1997) for examples of non-parametric matching.

exact matching is that conditioning on X_i eliminates selection bias, *i.e.*,

$$E[Y_{ti}|\ X_i, D_i = 1] = E[Y_{ti}|\ X_i, D_i = 0]; \quad t = 0, 1. \tag{1}$$

An implication of Rosenbaum and Rubin's propensity-score theorem is that if (1) is true, then it must also be true that conditioning on $p(X_i)$ eliminates selection bias,[2] *i.e.*,

$$E[Y_{ti}|\ p(X_i), D_i = 1] = E[Y_{ti}|\ p(X_i), D_i = 0]; \quad t = 0, 1. \tag{2}$$

This leads to an estimator of average treatment effects that has the following form:

$$\begin{aligned} E[Y_{1i} - Y_{0i}] &= E\{E[Y_{1i}|\ X_i, D_i = 1] - E[Y_{0i}|\ X_i, D_i = 0]\} \\ &= E\{E[Y_{1i}|\ p(X_i), D_i = 1] - E[Y_{0i}|\ p(X_i), D_i = 0]\}. \end{aligned} \tag{3}$$

The first line of expression (3) is the population analog of an exact-matching estimator, while the second line matches only on the propensity score. The value of propensity-score matching is in the dimension reduction generated by regions where $p(X_i)$ is constant while $E[Y_{1i}|\ X_i]$ or $E[Y_{0i}|\ X_i]$ are not constant. In a randomized trial, for example, $p(X_i)$ is constant, so there is no need to control for covariates to eliminate bias.

In practice, propensity-score matching is often based on an estimated propensity score as well as prior restrictions, such as a constant score or the elimination of certain covariates from the set of arguments of $p(X_i)$. Ultimately, of course, the use of an estimated propensity score in matching raises the same modeling issues as covariate-matching. However, applied researchers seem willing to make stronger assumptions about the propensity score than about the relationship between covariates and outcomes. A number of empirical examples using the propensity score suggest that this approach works reasonably well (see, *e.g.*, Rosenbaum and Rubin, 1984 and 1985a; Todd, 1995; Dehejia and Wahba, 1997; Imbens, Rubin, and Sacerdote, 1997; and Heckman, Ichimura, and Todd, 1998).

This evidence of practical utility notwithstanding, from a theoretical point of view, propensity-score-based estimators present some puzzles. Hahn (1998, Theorems 1 and 2) shows that for estimates of average treatment effects, the propensity score is ancillary, in the sense that knowledge of the propensity score does not lower the semiparametric efficiency bound. Moreover, covariate-matching is asymptotically efficient, *i.e.*, attains the semiparametric efficiency bound, while propensity-score matching does not. Finally, these theoretical results extend to the case where exact matching is not feasible, as with continuously distributed covariates, and the relevant conditional mean functions are instead approximated

[2] A premise of the propensity-score theorem and our discussion below is that (Y_{1i}, Y_{0i}) is independent of D_i given X_i, and not just mean independent.

using some kind of nonparametric regression (Hahn, 1998, Theorem 6). In short, asymptotic arguments would appear to offer no justification for anything other than full control for covariates.

The purpose of this paper is to develop formal theory and present some examples to substantiate the intuition that, because cell sizes may be small and some cells may be empty, there is a cost to covariate-matching, even if the covariates are discrete and exact matching is feasible. In some plausible scenarios, estimators that control only for the propensity score will in fact be more efficient (in finite samples) than exact-matching estimators, even though conventional asymptotic theory says otherwise. We make this point by first tabulating the finite-sample sampling variance of the covariate-matching and propensity-score-matching estimators for an analytically tractable special case. These results are also explained by analogy to well-known finite sample results for random effects panel-data models with non-stochastic regressors. We then present more general results based on an alternative asymptotic approximation where cell sizes are fixed but the number of cells becomes infinitely large. The asymptotic sequence used here is similar to sequences used by Bekker (1994), Bekker and van der Ploeg (1996) and Angrist and Krueger (1995) to analyze the finite-sample behavior of instrumental-variables estimators. The general results from this analysis show that propensity-score estimators can be more efficient than matching estimators when cell-sizes are small, the explanatory value of the covariates is low conditional on propensity score, and/or the probability of treatment is far from $\frac{1}{2}$.

The paper is organized as follows. In the next section, we outline the basic setup and compare the finite-sample behavior of two types of matching estimators in a simple model. Section 3 develops the panel-data version of the treatment effects problem, and an alternative asymptotic sequence based on increasing the number of cells of fixed size ("panel-asymptotics"). This section also discusses the possibility of producing a more efficient random-effects type estimator from a linear combination of covariate-matching and propensity-score-matching estimators. Section 4 discusses the likely generality of these results and presents some Monte Carlo evidence, which suggests that the new asymptotic sequence does indeed provide an accurate description of the relative finite-sample performance of matching and propensity-score estimators. Finally, Section 5 concludes and offers some directions for further work. All technical derivations are presented in an appendix.

2 Notation and Motivation

Throughout this paper, we assume that

Assumption 1 Treatment is ignorable (independent of potential outcomes) given covariates: $(Y_{0i}, Y_{1i}) \perp D_i \mid X_i$.

We further assume that

Assumption 2 X_i is multinomial, and takes K possible values, say $x^{(1)}, \ldots, x^{(K)}$, with probability $\frac{1}{K}$. We call $1\left(X_i = x^{(k)}\right)$ the k^{th} cell indicator.

This is a modelling device that allows us to change the number of cells. It is not really restrictive since $x^{(k)}$ can be anything. The multinomial assumption allows for a discrete approximation to any distribution for large K.[3]

We also assume that

Assumption 3 The propensity score $\Pr[D_i = 1 | X_i]$ is known to be fixed at π.

In models with discrete covariates, the difference between covariate-matching and propensity-score matching estimators arises from how the covariates are handled when the propensity score is constant. A fixed propensity score allows us to capture this idea very simply. Since observations are assumed to be independent across cells, the question of overall efficiency (*i.e.*, in a general setting with variable propensity score) is also addressed by looking at a single score-value. An analysis with known propensity score may provide an overly optimistic view of the relative performance of propensity-score-matching over covariate-matching, but since Hahn (1998) shows that knowledge of the propensity score does not affect the efficiency bound for average treatment effects, this seems to be a good starting point. Note also that estimated-propensity-score-matching is obviously equal to covariate-matching almost surely when X_i is discrete. This suggests that the notion of restrictions on the propensity score lies at the heart of the propensity-score/exact-matching distinction.

The two most commonly discussed parameters in evaluation studies are the effect of treatment on the treated $E[Y_{1i} - Y_{0i} | D_i = 1]$, and the average treatment effect $E[Y_{1i} - Y_{0i}]$. Since $\Pr[D_i = 1 | X_i] = \pi$ in our setup, $E[Y_{1i} - Y_{0i} | D_i = 1] = E[Y_{1i} - Y_{0i}]$. This equivalence allows us to sidestep the fact that knowledge of the propensity score can reduce the asymptotic variance bound for $E[Y_{1i} - Y_{0i} | D_i = 1]$, though this reduction does not come through matching on the propensity score. The propensity score is useful in this context because it weights covariate-specific comparisons underlying the effect-on-the-treated parameter.[4]

In most of the paper, we model cell size as fixed, so the sampling framework stratifies on X_i:

[3] See, *e.g.*, Chamberlain (1987) for a similar modeling strategy.

[4] While knowledge of the propensity score does reduce the semiparametric efficiency bound for the effect of treatment on the treated, the efficient estimator for this parameter still involves covariate-matching and not

Assumption 4 Each cell size is equal to M. We adopt the convention that the first n_{1k} individuals are treated in each cell, so that $n_{1k} \sim \text{Binomial}(M, \pi)$.

Stratified sampling is empirically relevant for some studies, but we adopt this assumption for technical reasons, since it simplifies the arguments and allows us to focus on the randomness in treatment status and outcomes within covariate-cells. It should also be noted that the traditional justification for the selection-on-observables assumption, $(Y_{0i}, Y_{1i}) \perp D_i \mid X_i$, makes X_i ancillary, so little would seem to be lost from stratification in this setting.

The following notation is useful:

Definition 1 *Let y_{0ki} and y_{1ki} denote potential outcomes under control and treatment for the i^{th} individual in the k^{th} cell. Let y_{0k} and y_{1k} denote the average potential outcomes under control and under treatment in the k^{th} cell. Also, let σ^2_{0k} and σ^2_{1k} denote the conditional variance of y_{0ki} and y_{1ki} in the k^{th} cell. Finally, let $\alpha_k \equiv y_{0k}$ and $\beta_k \equiv y_{1k} - y_{0k}$.*

We now define the two matching estimators considered in this paper. The covariate-matching estimator, b_c, is

$$b_c \equiv \frac{1}{\sum_{k=1}^{K} 1\left(1 \leq n_{1k} \leq M-1\right)} \sum_{k=1}^{K} 1\left(1 \leq n_{1k} \leq M-1\right) \left(\frac{1}{n_{1k}} \sum_{i=1}^{n_{1k}} y_{1ki} - \frac{1}{n_{0k}} \sum_{i=n_{1k}+1}^{M} y_{0ki} \right).$$

Because the propensity score is constant, matching on the propensity score is equivalent to ignoring covariates. The propensity-score-matching estimator, b_p, is therefore

$$b_p \equiv \frac{1}{\sum_{k=1}^{K} 1\left(1 \leq n_{1k}\right) n_{1k}} \sum_{k=1}^{K} \left[1\left(1 \leq n_{1k}\right) n_{1k} \left(\frac{1}{n_{1k}} \sum_{i=1}^{n_{1k}} y_{1ki} \right) \right]$$
$$- \frac{1}{\sum_{k=1}^{K} 1\left(n_{1k} \leq M-1\right) n_{0k}} \sum_{k=1}^{K} \left[1\left(n_{1k} \leq M-1\right) n_{0k} \left(\frac{1}{n_{0k}} \sum_{i=n_{1k}+1}^{M} y_{0ki} \right) \right].$$

There is some probability that matching on covariates and/or the propensity score cannot be implemented. For example, if all cells consist of only treated individuals, then matching on either covariates or the propensity score is infeasible. Note that both estimators are unbiased.[5]

matching on the propensity score (Hahn, Proposition 7). Heckman, Ichimura, and Todd's (1998) result that the propensity score does not necessarily lead to efficiency gains for estimates of the effect on the treated ignores the gain resulting from the fact that the propensity score is also the weighting function used to average covariate-specific contrasts.

[5]In practice, the question of (higher-order) bias is likely to arise in models with continuous covariates, where propensity-score-matching is based on an estimated propensity score, and covariate-matching is based

A simple example can be used to illustrate important differences in the finite sample behavior of b_c and b_p. Assume that

1. $K = 2$
2. The treatment effect is constant and equal to β_0, *i.e.*, $y_{0ki} = \alpha_k + \varepsilon_{ki}$, and $y_{1ki} = \beta_0 + \alpha_k + \varepsilon_{ki}$.
3. $E[\varepsilon_{ki}^2] = 1$. (a normalization)

In order to understand the finite sample behavior of the two estimators, we consider three cases. In Case 1, both cells contain treated and control observations. In Case 2, one of the two cells consists of treated or controls only. In Case 3, each cell consists of treated or controls only, so b_c cannot be computed. We therefore focus on variance comparisons conditional on the event that b_c exists. The efficiency of b_p relative to b_c is defined to be $\sqrt{\text{Var}(b_c)} \Big/ \sqrt{\text{Var}(b_p)}$, where $\text{Var}(b_c)$ and $\text{Var}(b_p)$ denote the *conditional* variance of b_c and b_p given the event that both are computable. Without loss of generality, we *define*

$$\text{Var}(\alpha_k) \equiv \frac{1}{2}\sum_{k=1}^{2}(\alpha_k - \overline{\alpha})^2, \quad \text{where} \quad \overline{\alpha} \equiv \frac{1}{2}\sum_{k=1}^{2}\alpha_k. \tag{4}$$

Note that the R^2 in the theoretical regression of the outcomes on covariates (cell indicators) can be written $\text{Var}(\alpha_k)/(\text{Var}(\alpha_k) + 1)$. We consider the relative efficiency of b_p for various (π, M, R^2) combinations. Tables 1 and 2 report the relative efficiency of b_p measured by the ratio $\sqrt{\text{Var}(b_c)} \Big/ \sqrt{\text{Var}(b_p)}$ for $\pi = .1$ and $.5$. This is an exact finite sample calculation, the details of which are discussed in Appendix A. The tables show that the relative efficiency of b_p increases as R^2 falls, M falls, and π falls, and that b_p is actually more efficient than b_c for some (π, M, R^2) combinations. These exact calculations suggest that conventional (large cell) asymptotic approximations may provide a poor guide to the relative precision of these estimators in some applications.

We also ask whether the relative efficiency of b_p for some (π, M, R^2) combinations is solely a consequence of the fact that b_p typically uses more observations than b_c. Consider the relative efficiency in Case 1, where both estimators use the same number of observations.

on non-parametric estimates of the conditional mean function of outcomes given covariates. Our panel-asymptotic framework is limited to discrete covariates and situations where the researcher is willing to assume that the propensity score has known "flat spots". An analysis of other cases raises statistical and modeling issues beyond the scope of this paper. See Heckman, Ichimura, and Todd (1998) for an analysis with continuous X_i and estimated $p(X_i)$ under (first-order) conventional asymptotics.

Any difference in variance in this case can therefore be attributed to the efficiency with which each estimator processes information. When $M = 2$, $b_p = b_c$ in Case 1, so obviously there is no efficiency difference. The relative variance in Case 1 when $M = 3$ is

$$\sqrt{\frac{27}{(32\pi - 32\pi^2)\frac{R^2}{1-R^2} + 27 - 6\pi + 6\pi^2}}.$$

When $\pi = .5$, this equals

$$5.196\,2\sqrt{\left(\frac{1}{8\frac{R^2}{1-R^2} + 25.5}\right)},$$

which shows b_p to be moderately more efficient for $R^2 \leq .157\,89$. Thus, the relative finite-sample efficiency of b_p arises not only because b_p uses more observations; the potential benefit from propensity-score-matching apparently comes partly from the fact that cell-specific contrasts can be imprecise.

3 Panel Characterization

We now introduce a panel characterization of the evaluation model in the previous section. The panel analogy enables us to draw on the econometric literature dealing with problems of this type. See, for example, Wallace and Hussain (1969), Maddala (1971), Chamberlain and Griliches (1975), Mundlak (1978), Hausman and Taylor (1981), and Chamberlain (1984). We argue that covariate-matching and propensity-score-matching estimators are within-type and pooled estimators. Standard results for random-effects panel models with nonstochastic regressors suggest that neither within-estimators nor pooled estimators are efficient, and that their precision cannot be ranked unambiguously.

3.1 Random Coefficient Model

The panel equivalent of the evaluation problem looks like this. Let D_{ki} denote a binary treatment indicator, and write the observed y_{ki} as

$$\begin{aligned} y_{ki} &\equiv D_{ki}y_{1ki} + (1 - D_{ki})\, y_{0ki} = y_{0ki} + (y_{1ki} - y_{0ki})\, D_{ki} \\ &= \alpha_k + \beta_k D_{ki} + \{(y_{0ki} - \alpha_k) + (y_{1ki} - y_{0ki} - \beta_k)\, D_{ki}\}. \end{aligned}$$

With ε_{ki} defined as the residual in the above equation, we can write

$$y_{ki} = \alpha_k + \beta_k D_{ki} + \varepsilon_{ki}, \quad k = 1, \ldots, K; i = 1, \ldots, M \tag{5}$$

where the parameter of interest is equal to $E[\beta_k]$. This is the random-coefficient panel model considered by Swamy (1970), and Chamberlain (1992).

Consider first the simple model where β_k is fixed at β_0 and ε_{ki} is homoscedastic with variance σ_ε^2. Observe that α_k is independent of D_{ki} because of Assumption 1, so $\text{Var}(\alpha_k | D_{ki}) = \text{Var}(\alpha_k) \equiv \sigma_\alpha^2$. Conditional on the realization of D_{ki} and assuming that both b_p and b_c can be computed, it is easy to see that, under these assumptions, (5) is the traditional random-effects panel model with nonstochastic regressors

$$y_{ki} = \alpha_k + \beta_0 D_{ki} + \varepsilon_{ki}. \tag{6}$$

The efficient unbiased estimator for this model is well-known to be a weighted average of between and within estimators, or equivalently, within and pooled estimators.[6] Recall that

$$b_c = \frac{1}{K}\sum_{k=1}^{K} \widehat{b}_k,$$

where $\widehat{b}_k$ is the OLS estimator of β_k for the k^{th} cell. Because b_c is the sample average of estimators using only within-cell variation, b_c is also a within-type estimator, though it is not equal to the traditional within estimator, which can be written in this case as

$$b_w = \sum_{k=1}^{K} \widehat{\pi}_k (1 - \widehat{\pi}_k) \widehat{b}_k \Big/ \sum_{k=1}^{K} \widehat{\pi}_k (1 - \widehat{\pi}_k),$$

where $\widehat{\pi}_k = \frac{1}{K}\sum_{i=1}^{M} D_{ki}$.[7] Note also that b_p is the OLS coefficient from a regression of y on D, and is therefore the traditional pooled estimator that ignores group structure. This suggests that b_c and b_p cannot be ranked unambiguously, since neither within-type nor pooled estimators are efficient for random-effects panel models.

To substantiate this conjecture, we calculate the finite sample variances of b_c and b_p treating D_{ki} as nonstochastic, α_k as random, assuming constant treatment effects and homoscedastic errors, and conditional on b_c and b_p both being computable. In this setting, it can be shown that

$$\text{Var}(b_c) = \sigma_\varepsilon^2 \cdot \frac{1}{K^2}\sum_{k=1}^{K}\left(\frac{1}{n_{1k}} + \frac{1}{M - n_{1k}}\right),$$

[6] See, *e.g.*, Maddala (1971).

[7] See, *e.g.*, Angrist (1998).

and

$$\text{Var}\,(b_p) = \sigma_\alpha^2 \cdot \sum_{k=1}^{K} \left(\frac{n_{1k}}{\sum_{k=1}^{K} n_{1k}} - \frac{M - n_{1k}}{KM - \sum_{k=1}^{K} n_{1k}} \right)^2 + \sigma_\varepsilon^2 \cdot \left(\frac{1}{\sum_{k=1}^{K} n_{1k}} + \frac{1}{KM - \sum_{k=1}^{K} n_{1k}} \right).$$

For example, if $K = 2$, $M = 3$, $n_{11} = 1$, and $n_{12} = 2$, we have

$$\text{Var}\,(b_c) = \frac{3}{4}\sigma_\varepsilon^2, \quad \text{and} \quad \text{Var}\,(b_p) = \frac{2}{9}\sigma_\alpha^2 + \frac{2}{3}\sigma_\varepsilon^2.$$

Hence, the difference between them is

$$\text{Var}\,(b_c) - \text{Var}\,(b_p) = \frac{1}{12}\sigma_\varepsilon^2 - \frac{2}{9}\sigma_\alpha^2,$$

which is of ambiguous sign.[8]

How can this ambiguity be reconciled with Hahn's (1998) result that b_c is asymptotically more efficient than b_p? Traditional asymptotic arguments fix the data generating process, and let the number of observations grow to infinity. In our setting, this asymptotic sequence would have K fixed while $M \to \infty$, and the between-cell variation would become noninformative as the sample size increases. In fact, it is well-known that the random effects estimator converges to the fixed effects (matching) estimator under this asymptotic sequence, so there is no contradiction. On the other hand, for small M, a panel-type asymptotic sequence with M fixed while $K \to \infty$ may be more appropriate. We turn to this question in the next section.

3.2 Panel Asymptotics

Rosenbaum and Rubin's principal motivation for propensity score matching seems to have been the possibility that some cells have to be dropped in procedures based on covariate-matching.[9] The two-cell example in Section 2 shows that propensity-score matching does more than increasing the number of observations used in estimation, however. To provide a general statement of this result, we use an alternative asymptotic approximation where cell sizes are fixed and the number of cells grows to infinity. As noted above, this corresponds to the usual large-cross-section/small-time-series asymptotic approximation for panel data.

[8] Here, the α_k are assumed to be random variables, whereas the α_k were treated as fixed constants in the previous section.

[9] See, *e.g.*, Rosenbaum and Rubin (1984, p. 516).

The analog of the cross-section dimension in our case is K, and the analog of the time-series dimension is M.

As a regularity condition, we assume that

Assumption 5 The sequence $\{(y_{0k1}, y_{1k1}, \ldots, y_{0kM}, y_{1kM}), k = 1, 2, \ldots\}$ is *i.i.d.* Furthermore, for given k, (y_{1k1}, y_{0k1}), $\ldots$, (y_{1kM}, y_{0kM}) are *i.i.d.*

Obviously, Assumption 5 implies that, in the panel characterization, (5),

$$(\alpha_k, \beta_k, D_{k1}, \ldots, D_{kM}, \varepsilon_{k1}, \ldots, \varepsilon_{kM})$$

is *i.i.d.* Note that this means we approximate sampling distributions without assuming any prior information on α_k, and β_k. This is consistent with the nonparametric spirit of matching procedures. As before, our objective is to estimate the average treatment effect, $\beta \equiv E[y_{1ki} - y_{0ki}]$. The main theoretical result is given below:

Theorem 1 *Under Assumptions 1 - 5, we have*

$$\sqrt{K}(b_c - \beta) \to \mathcal{N}(0, \omega_c^2), \qquad \sqrt{K}(b_p - \beta) \to \mathcal{N}(0, \omega_p^2),$$

where

$$\omega_c^2 \equiv \frac{g(\pi, M) E[\sigma_{1k}^2] + g(1-\pi, M) E[\sigma_{0k}^2]}{\left(1 - \pi^M - (1-\pi)^M\right)^2} + \frac{\mathrm{Var}(y_{1k} - y_{0k})}{1 - \pi^M - (1-\pi)^M},$$

and

$$\omega_p^2 \equiv \frac{1}{M\pi} E[\sigma_{1k}^2] + \frac{1}{M(1-\pi)} E[\sigma_{1k}^2] + \mathrm{Var}(y_{1k} - y_{0k}) + \frac{1}{M} \mathrm{Var}\left(\sqrt{\frac{1-\pi}{\pi}} y_{1k} + \sqrt{\frac{\pi}{1-\pi}} y_{0k}\right),$$

where

$$g(\pi, M) \equiv \sum_{k=1}^{M-1} \binom{M}{k} \pi^k (1-\pi)^{M-k} \frac{1}{k}.$$

Proof. See Appendix B. ■

The implications of this result for the relative sampling variance of b_c and b_p can be summarized using $\sqrt{\omega_c^2/\omega_p^2}$. This expression is complicated but can be tabulated, or simplified for special cases. Tables 3 and 4 report the relative efficiency of b_p for two values

of π assuming a constant treatment effect and homoscedastic errors. As before, we define the theoretical R^2 as $\sigma_\alpha^2/(\sigma_\alpha^2+\sigma_\varepsilon^2)$. Note that b_p is predicted to be more efficient if and only if $\sqrt{\omega_c^2/\omega_p^2}>1$. As before, these tables show that the relative efficiency of b_p typically increases as R^2 falls, M falls, and π falls, and that b_p is actually more efficient than b_c for some (π, M, R^2) combinations.

In Section 4, we turn to the question of whether Theorem 1 captures the actual finite sample behavior of b_c and b_p in samples with random cell sizes.

3.3 Comparison with Conventional Asymptotics

How do panel-asymptotic results differ from conventional asymptotic results, where the number of cells is fixed and cell sizes are random and increasing? Let $\mathbb{N}$ and M^* denote total sample size and average cell size in a random sample. Using an $\mathbb{N}\to\infty$ conventional asymptotic sequence, where M^* grows to ∞ as a consequence while K is fixed, we can show[10] that

$$\sqrt{\mathbb{N}}\left(b_c-\beta\right)\to\mathcal{N}\left(0,\frac{E\left[\sigma_{1k}^2\right]}{\pi}+\frac{E\left[\sigma_{0k}^2\right]}{1-\pi}+\mathrm{Var}\left(y_{1k}-y_{0k}\right)\right),\quad\text{and}$$

$$\sqrt{\mathbb{N}}\left(b_p-\beta\right)\to\mathcal{N}\left(0,\frac{E\left[\sigma_{1k}^2\right]}{\pi}+\frac{E\left[\sigma_{0k}^2\right]}{1-\pi}+\mathrm{Var}\left(y_{1k}-y_{0k}\right)+\mathrm{Var}\left(\sqrt{\frac{1-\pi}{\pi}}y_{1k}+\sqrt{\frac{\pi}{1-\pi}}y_{0k}\right)\right).$$

So conventional asymptotics approximates finite sample variances as:

$$\mathrm{CVar}\left(b_c\right)\equiv\frac{1}{\mathbb{N}}\left(\frac{E\left[\sigma_{1k}^2\right]}{\pi}+\frac{E\left[\sigma_{0k}^2\right]}{1-\pi}+\mathrm{Var}\left(y_{1k}-y_{0k}\right)\right), \tag{7}$$

and

$$\mathrm{CVar}\left(b_p\right)\equiv\frac{1}{\mathbb{N}}\left(\frac{E\left[\sigma_{1k}^2\right]}{\pi}+\frac{E\left[\sigma_{0k}^2\right]}{1-\pi}+\mathrm{Var}\left(y_{1k}-y_{0k}\right)+\mathrm{Var}\left(\sqrt{\frac{1-\pi}{\pi}}y_{1k}+\sqrt{\frac{\pi}{1-\pi}}y_{0k}\right)\right). \tag{8}$$

The last term in $\mathrm{CVar}(b_p)$ can be interpreted as the penalty for failure to control for covariates under conventional asymptotics. Note that, with constant treatment effects, this term equals zero if the between cell variance is zero.

Panel-asymptotics approximates finite sample variances as:

$$\mathrm{PVar}\left(b_c\right)\equiv\frac{1}{\mathbb{N}}\left(\frac{Mg\left(\pi,M\right)E\left[\sigma_{1k}^2\right]+Mg\left(1-\pi,M\right)E\left[\sigma_{0k}^2\right]}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}+\frac{M\,\mathrm{Var}\left(y_{1k}-y_{0k}\right)}{1-\pi^M-\left(1-\pi\right)^M}\right), \tag{9}$$

[10] See Hahn (1998).

and

$$\mathrm{PVar}\left(b_p\right) \equiv \frac{1}{\mathbb{N}}\left(\frac{1}{\pi}E\left[\sigma_{1k}^2\right] + \frac{1}{1-\pi}E\left[\sigma_{1k}^2\right] + M\,\mathrm{Var}\left(y_{1k}-y_{0k}\right) + \mathrm{Var}\left(\sqrt{\frac{1-\pi}{\pi}}y_{1k} + \sqrt{\frac{\pi}{1-\pi}}y_{0k}\right)\right), \tag{10}$$

where we used the fact that $\mathbb{N} = KM$ in the panel-asymptotic sequence. The penalty term in $\mathrm{CVar}(b_p)$ remains in $\mathrm{PVar}(b_p)$, but now the terms

$$\frac{E\left[\sigma_{1k}^2\right]}{\pi} \quad \text{and} \quad \frac{E\left[\sigma_{0k}^2\right]}{1-\pi}$$

in $\mathrm{CVar}(b_c)$ become

$$\frac{Mg\left(\pi, M\right)E\left[\sigma_{1k}^2\right]}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2} \quad \text{and} \quad \frac{Mg\left(1-\pi, M\right)E\left[\sigma_{0k}^2\right]}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}$$

under panel-asymptotics in (9). This is because the first two terms partly reflect the fact that some cells may have to be dropped in the computation of b_c. Note also that the panel-asymptotic approximation inflates the third term in $\mathrm{CVar}(b_c)$ and $\mathrm{CVar}(b_p)$, which is

$$\mathrm{Var}\left(y_{1k}-y_{0k}\right)$$

in both expressions. This term becomes

$$\frac{M\,\mathrm{Var}\left(y_{1k}-y_{0k}\right)}{1-\pi^M-\left(1-\pi\right)^M} \quad \text{and} \quad M\,\mathrm{Var}\left(y_{1k}-y_{0k}\right)$$

in (9) and (10). The inflation factor, $\frac{M}{1-\pi^M-(1-\pi)^M}$ in (9), is larger than M in (10). This partly reflects the fact that the conventional asymptotic approximation is more optimistic about the precision with which realized cell-differences are actually estimated. Note also that the inflation factor is larger for π close to zero or one.

To summarize the difference between the two approximations, we write

$$\begin{aligned}\left(\mathrm{PVar}\left(b_c\right)-\mathrm{PVar}\left(b_p\right)\right)-\left(\mathrm{CVar}\left(b_c\right)-\mathrm{CVar}\left(b_p\right)\right) &= \frac{1}{\mathbb{N}}\left(\frac{Mg\left(\pi, M\right)}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}-\frac{1}{\pi}\right)E\left[\sigma_{1k}^2\right] \\ &+ \frac{1}{\mathbb{N}}\left(\frac{Mg\left(1-\pi, M\right)}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}-\frac{1}{1-\pi}\right)E\left[\sigma_{0k}^2\right] \\ &+ \frac{1}{\mathbb{N}}\left(\frac{M}{1-\pi^M-\left(1-\pi\right)^M}-M\right)\mathrm{Var}\left(y_{1k}-y_{0k}\right).\end{aligned}$$

The first two terms on the right are nonnegative.[11] Note that the third term is zero if and only if $\text{Var}\left(y_{1k} - y_{0k}\right) = 0$. We therefore expect the finite sample advantage of b_p to be larger with heterogeneous treatment effects.

3.4 Linear Combinations of b_c and b_p

Since neither b_c nor b_p is efficient, we now ask whether we can construct a treatment-effects estimator that is more efficient than both. In this case, a more efficient estimator can be obtained from the minimum variance linear combination of b_c and b_p:[12]

$$b^* = \xi b_c + (1 - \xi) b_p.$$

The asymptotic variance of b^* is minimized by choosing

$$\xi = \frac{\text{Var}_a\left(b_p\right) - \text{Cov}_a\left(b_c, b_p\right)}{\text{Var}_a\left(b_c\right) - 2\,\text{Cov}_a\left(b_c, b_p\right) + \text{Var}_a\left(b_p\right)},$$

where Var_a and Cov_a denote asymptotic variance and asymptotic covariance under the asymptotic sequence in Theorem 1. The variance terms are available from Theorem 1.

[11] It can be shown that

$$\frac{Mg\left(\pi, M\right)}{\left(1 - \pi^M - (1-\pi)^M\right)^2} \geq \frac{1}{\pi - \pi^M}.$$

Let Z have the same distribution as $n_{1k} \sim \text{Binomial}\left(M, \pi\right)$ conditional on the event that $1 \leq n_{1k} \leq M - 1$. We then have

$$E\left[Z\right] = \frac{\sum_{k=1}^{M-1} \binom{M}{k} \pi^k (1-\pi)^{M-k} k}{1 - \pi^M - (1-\pi)^M} = \frac{M\pi - M\pi^M}{1 - \pi^M - (1-\pi)^M}$$

and

$$E\left[\frac{1}{Z}\right] = \frac{\sum_{k=1}^{M-1} \binom{M}{k} \pi^k (1-\pi)^{M-k} \frac{1}{k}}{1 - \pi^M - (1-\pi)^M} = \frac{g\left(\pi, M\right)}{1 - \pi^M - (1-\pi)^M}$$

By Jensen's Inequality, this means

$$\frac{g\left(\pi, M\right)}{1 - \pi^M - (1-\pi)^M} = E\left[\frac{1}{Z}\right] \geq \frac{1}{E\left[Z\right]} = \frac{1 - \pi^M - (1-\pi)^M}{M\pi - M\pi^M},$$

from which the desired relationship follows.

[12] Unlike in the traditional random effects model, we cannot claim that the resulting estimator is actually optimal since we have not derived the semiparametric efficiency bound for the sort of panel model considered here. Chamberlain (1992, Section 4) presents a bound for such a model, but his bound does not impose independence of $\varepsilon_{ki}, i = 1, \ldots, M$.

The covariance term is

$$\frac{1}{\pi M}E\left[\sigma_{1k}^2\right]+\frac{1}{M\left(1-\pi\right)}E\left[\sigma_{0k}^2\right]$$
$$+\frac{1-\pi^{M-1}}{1-\pi^M-\left(1-\pi\right)^M}\operatorname{Var}\left(y_{1k}\right)+\frac{1-\left(1-\pi\right)^{M-1}}{1-\pi^M-\left(1-\pi\right)^M}\operatorname{Var}\left(y_{0k}\right)$$
$$+\frac{-2+\pi^{M-1}+\left(1-\pi\right)^{M-1}}{1-\pi^M-\left(1-\pi\right)^M}\operatorname{Cov}\left(y_{1k},y_{0k}\right).$$

See Appendix B for the derivation of this formula. Note that as in Theorem 1, b^* was computed for M fixed.[13]

To develop intuition for the weighting formula, suppose as before that (i) the treatment effect is constant and equal to β_0, i.e., $y_{0ki}=\alpha_k+\varepsilon_{ki}$, and $y_{1ki}=\beta_0+\alpha_k+\varepsilon_{ki}$; (ii) ε_{ki} has variance equal to σ_ε^2; and (iii) α_k has mean μ_α, and variance σ_α^2. After some algebra, we obtain following simplification:

$$\operatorname{Var}_a\left(b_p\right)-\operatorname{Cov}_a\left(b_c,b_p\right)=\frac{1}{M\pi\left(1-\pi\right)}\sigma_\alpha^2,$$

and

$$\operatorname{Var}_a\left(b_c\right)-2\operatorname{Cov}_a\left(b_c,b_p\right)+\operatorname{Var}_a\left(b_p\right)$$
$$=\left(\frac{g\left(\pi,M\right)}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}+\frac{g\left(1-\pi,M\right)}{\left(1-\pi^M-\left(1-\pi\right)^M\right)^2}-\frac{1}{\pi M}-\frac{1}{M\left(1-\pi\right)}\right)\sigma_\varepsilon^2$$
$$+\frac{1}{M\pi\left(1-\pi\right)}\sigma_\alpha^2.$$

Therefore, the optimal weight is equal to

$$\xi^*=\frac{\frac{1}{\pi(1-\pi)}\sigma_\alpha^2}{\left(\frac{Mg(\pi,M)}{\left(1-\pi^M-(1-\pi)^M\right)^2}+\frac{Mg(1-\pi,M)}{\left(1-\pi^M-(1-\pi)^M\right)^2}-\frac{1}{\pi}-\frac{1}{(1-\pi)}\right)\sigma_\varepsilon^2+\frac{1}{\pi(1-\pi)}\sigma_\alpha^2}.$$

Without loss of generality, we may normalize $\sigma_\varepsilon^2=1$. Note that (i) $\xi^*\to 1$ as $M\to\infty$ ($b_c\approx b^*$) ; and (ii) $\xi^*\to 0$ as $\sigma_\alpha^2\to 0$ ($b_p\approx b^*$). In other words, the linear-combination estimator converges to the covariate-matching estimator as the cell size gets large and/or the between-cell variance gets small. On the other hand, the linear-combination estimator

[13]In the next section we show that Theorem 1 provides a reasonable approximation to finite-sample behavior even when cell sizes are random. But we leave the development of an efficient estimator for this case for future work.

converges to the propensity-score-matching estimator as the random effects variance gets small. This is analogous to the behavior of the random effects GLS estimator for traditional panel models with constant treatment effects: GLS converges to the within-estimator as the time series dimension gets large and/or the variance of the random individual effects gets small, while convergence to the pooled estimator occurs in the opposite case.[14]

4 Validity of the Approximation

The panel-asymptotic results in Theorem 1 were derived under stratified sampling, for covariate cells of fixed size. Much of the discussion also relied on the simplifying assumption of constant treatment effects. In this section, we compare the finite-sample behavior predicted by Theorem 1 with actual finite sample behavior under random sampling. We begin with constant treatment effects, because this assumption allows an analytic derivation of the finite-sample variance conditional on cell sizes. We then do a Monte Carlo integration to allow for random cell sizes. Finally, we report the results from Monte Carlo experiments with heterogeneous treatment effects.

Tables 5 and 6 report actual finite sample behavior under random sampling with constant treatment effects. We again consider a model where $y_{0ki} = \alpha_k + \varepsilon_{ki}$, and $y_{1ki} = \beta_0 + \alpha_k + \varepsilon_{ki}$, with $\text{Var}(\varepsilon_{ki}) = \sigma_\varepsilon^2$ and $\text{Var}(\alpha_k) = \sigma_\alpha^2$. We assume that there are K cells, with cell sizes equal to $M_1, \ldots, M_K$. Let M^* denote the average cell size. Note that

$$b_c = \beta_0 + \frac{1}{\sum_{k=1}^K M_k 1(1 \leq n_{1k} \leq M_k - 1)} \sum_{k=1}^K M_k 1(1 \leq n_{1k} \leq M_k - 1) \left(\frac{1}{n_{1k}} \sum_{i=1}^{n_k} \varepsilon_{ki} - \frac{1}{M_k - n_{1k}} \sum_{i=n_k+1}^{M_k} \varepsilon_{ki} \right),$$

with conditional variance given $(M_1, \ldots, M_K, n_{11}, \ldots, n_{1K})$ equal to

$$\sigma_\varepsilon^2 \sum_{k=1}^K \left(\frac{M_k}{\sum_{k=1}^K M_k 1(1 \leq n_{1k} \leq M_k - 1)} \right)^2 \left(\frac{1}{n_{1k}} + \frac{1}{M_k - n_{1k}} \right) 1(1 \leq n_{1k} \leq M_k - 1). \quad (11)$$

[14] Swamy (1970) derives the maximum likelihood estimator of β assuming normality of $(\alpha_k, \beta_k, \varepsilon_{ki})$, known error variances, and non-stochastic regressors (D_{ki}). This estimator is efficient under panel asymptotics if the error variances are common across cells. Except under constant treatment effects, the Swamy estimator does not appear to simplify to a linear combination of b_p and b_c.

Also, note that

$$b_p = \beta_0 + \sum_{k=1}^{K} \left(\frac{n_{1k}}{\sum_{k=1}^{K} n_{1k}} - \frac{M_k - n_{1k}}{\sum_{k=1}^{K} (M_k - n_{1k})} \right) \alpha_k + \frac{1}{\sum_{k=1}^{K} n_{1k}} \sum_{k=1}^{K} \sum_{i=1}^{n_{1k}} \varepsilon_{ki} - \frac{1}{\sum_{k=1}^{K} (M_k - n_{1k})} \sum_{k=1}^{K} \sum_{i=n_{1k}+1}^{M_k} \varepsilon_{ki},$$

with conditional variance equal to

$$\sigma_\alpha^2 \sum_{k=1}^{K} \left[\left(\frac{n_{1k}}{\sum_{k=1}^{K} n_{1k}} - \frac{M_k - n_{1k}}{\sum_{k=1}^{K} (M_k - n_{1k})} \right) \right]^2 + \sigma_\varepsilon^2 \left(\frac{1}{\sum_{k=1}^{K} n_{1k}} + \frac{1}{\sum_{k=1}^{K} (M_k - n_{1k})} \right). \quad (12)$$

We set $K = 100$, and assume that $(M_1, \ldots, M_K)$ are generated by a multinomial distribution with equal weights.[15] Results using 300 replications of $(M_1, \ldots, M_K, n_{11}, \ldots, n_{1K})$ to integrate (11) and (12) are reported in Tables 5 and 6. As before, we observe that the relative efficiency of b_p, measured by $\frac{\sqrt{\text{Var}(b_c)}}{\sqrt{\text{Var}(b_p)}}$, typically increases as the R^2 falls, M^* falls, and π falls, and that b_p is actually more efficient than b_c for some (π, M^*, R^2) combinations. The relative efficiency calculated allowing for random cell sizes is remarkably close to the ratio in Tables 3 and 4, calculated using our panel-asymptotic sequence.

We also conducted a small Monte Carlo study of a model with heterogeneous treatment effects. Figures 1 and 2 compare Monte Carlo sampling distributions under random sampling with heterogeneous treatment effects to the corresponding panel-asymptotic approximation. Again, we consider a model where $y_{0ki} = \alpha_k + \varepsilon_{ki}$, and $y_{1ki} = \beta_k + \alpha_k + \varepsilon_{ki}$, with $\text{Var}(\varepsilon_{ki}) = \sigma_\varepsilon^2$ and $\text{Var}(\alpha_k) = \sigma_\alpha^2$. Both figures set $K = 30$, the covariate $R^2 = .1$, and the propensity score $=.1$. We used 500 Monte Carlo replications. Figure 1 shows results from a model where $\beta_k \sim \text{Binomial}\left(1, \frac{1}{2}\right)$ independent of α_k and ε_{ki}. Figure 2 shows results from a model where $\alpha_k \sim \mathcal{N}(\alpha, \sigma_\alpha^2)$ and $\beta_k = 1(\alpha_k < \alpha)$. In this case, treatment effects are negatively correlated with untreated outcomes. The panel-asymptotic approximation predicts the Monte Carlo efficiency ratio reasonably well in both figures.

5 Conclusions and Directions for Future Work

Asymptotic theory provides a powerful and flexible tool for the analysis of the theoretical properties of alternative estimators, but empirical researchers and econometricians have become increasingly aware that conventional asymptotic results can be misleading. Recent

[15] We fix the total sample size, then break the sample up into N subsamples with expected size $\frac{\sum_{k=1}^{K} M_k}{K}$. This is equivalent to random sampling from a multinomial distribution where $\Pr\left(X_i = x^{(k)}\right) = p$ for all j.

examples of analyses using alternative approximations include the Chamberlain and Imbens (1996) and Staiger and Stock (1997) discussions of the bias in two-stage least squares, and the analyses of instrumental variables and grouping estimators by Bekker (1994) and Bekker and van der Ploeg (1996).

In this paper, we show that conventional asymptotic arguments may also be misleading for efficiency comparisons. The particular problem of concern here is whether to control for covariates in the estimation of treatment effects. Not surprisingly, in many cases that seem likely to be of practical importance, matching on the propensity score, which suffices to eliminate bias, is also more efficient than full covariate-matching. The results presented here, based on an analogy with random-effects models for panel data, provide some general guidelines for when this is most likely to be true. In future work, we hope to make these guidelines more specific, and to develop sharper results on an efficiency bound for random-effects estimators of the type introduced here.

Appendix

A Finite Sample Variance in 2-Cell Example

This calculation begins with the bias and variance of the two estimators, conditional on n_{11} and n_{12}, for cases where both estimators are defined. b_p is conditionally biased, though b_c is not.

A.1 b_c

Case 1 In this case, both cells contain treated and control observations. The conditional distribution of b_c given (n_{11}, n_{12}) has bias 0 and variance $\frac{1}{4}\left(\frac{1}{n_{11}} + \frac{1}{M-n_{11}} + \frac{1}{n_{12}} + \frac{1}{M-n_{12}}\right)$. Therefore, the conditional mean squared error is

$$w_1(n_{11}, n_{12}) \equiv \frac{1}{4}\left(\frac{1}{n_{11}} + \frac{1}{M-n_{11}} + \frac{1}{n_{12}} + \frac{1}{M-n_{12}}\right).$$

Case 2 In this case, either one of the two cells consists of treated or controls only. If the first cell is discarded but the second one is not, the conditional distribution of b_c is such that the bias is 0 and the variance is equal to $\frac{1}{n_{12}} + \frac{1}{M-n_{12}}$. Therefore, the conditional mean squared error is

$$w_2(n_{12}) \equiv \frac{1}{n_{12}} + \frac{1}{M-n_{12}}.$$

Similar comments apply when the second cell is discarded.

Case 3 In this case, each cell consists of treated or controls only. Since this happens in each cell with probability $\Pr(n_{1k} = 0 \text{ or } M) = \pi^M + (1-\pi)^M$, with probability $\left(\pi^M + (1-\pi)^M\right)^2$, the covariate-matching estimator is undefined.

Now, integrate over the distribution of n_{11}, and n_{12} using the fact that they are independent Binomial (M, π) random variables:

$$\sum_{n_{11}=1}^{M-1}\sum_{n_{12}=1}^{M-1} w_1(n_{11}, n_{12}) \binom{M}{n_{11}} \pi^{n_{11}} (1-\pi)^{M-n_{11}} \binom{M}{n_{12}} \pi^{n_{12}} (1-\pi)^{M-n_{12}}$$
$$+ \left(\pi^M + (1-\pi)^M\right) \sum_{n_{12}=1}^{M-1} w_2(n_{12}) \binom{M}{n_{12}} \pi^{n_{12}} (1-\pi)^{M-n_{12}}$$
$$+ \left(\pi^M + (1-\pi)^M\right) \sum_{n_{11}=1}^{M-1} w_2(n_{11}) \binom{M}{n_{11}} \pi^{n_{11}} (1-\pi)^{M-n_{11}}.$$

Dividing the above expression by $1 - \left(\pi^M + (1-\pi)^M\right)^2$, we obtain the variance of interest.

A.2 b_p

We consider the finite sample distribution of b_p for cases where b_c can be computed.[16]

Case 1 In this case,

$$b_p = \frac{1}{\sum_{k=1}^{2} n_{1k}} \sum_{k=1}^{2}\sum_{i=1}^{n_{1k}} (\alpha_k + \beta + \varepsilon_{ki}) - \frac{1}{2M - \sum_{k=1}^{2} n_{1k}} \sum_{k=1}^{2}\sum_{i=n_{1k}+1}^{M} (\alpha_k + \varepsilon_{ki})$$

with conditional bias $\frac{M(n_{11}-n_{12})}{(n_{11}+n_{12})(2M-n_{11}-n_{12})}(\alpha_1 - \alpha_2)$, and variance $\frac{2M}{(n_{11}+n_{12})(2M-(n_{11}+n_{12}))}$. Therefore, the conditional mean squared error is given by

$$\left(\frac{M(n_{11}-n_{12})}{(n_{11}+n_{12})(2M-n_{11}-n_{12})}(\alpha_1-\alpha_2)\right)^2 + \frac{2M}{(n_{11}+n_{12})(2M-(n_{11}+n_{12}))}.$$

Case 2 (i) Consider the case where the first cell consists of all treated, but the second cell is not dropped by b_c. We then have

$$b_p = \frac{1}{M+n_{12}}\left(\sum_{i=1}^{M}(\alpha_1+\beta_0+\varepsilon_{1i}) + \sum_{i=1}^{n_{12}}(\alpha_2+\beta+\varepsilon_{2i})\right) - \frac{1}{M-n_{12}}\sum_{i=n_{12}+1}^{M}(\alpha_2+\varepsilon_{2i})$$

which has the conditional bias $\frac{M}{M+n_{12}}(\alpha_1-\alpha_2)$, and variance $\frac{1}{M+n_{12}}+\frac{1}{M-n_{12}}$. Therefore, the conditional mean squared error is given by

$$\left(\frac{M}{M+n_{12}}(\alpha_1-\alpha_2)\right)^2 + \frac{1}{M+n_{12}} + \frac{1}{M-n_{12}}.$$

Similar comments apply when the second cell consists of all treated.

[16] Technically speaking, we may be able to define propensity score estimator even for case 3. We can have the first cell consisting of all treated and the second cell all consisting of control.

Casd 2 (ii) Consider the case where the first cell consists of all controls, but the second cell is not dropped by the covariate estimator. We then have

$$b_p = \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} (\alpha_2 + \beta_0 + \varepsilon_{2i}) - \frac{1}{2M - n_{12}} \left(\sum_{i=1}^{M} (\alpha_1 + \varepsilon_{1i}) + \sum_{i=n_{12}+1}^{M} (\alpha_2 + \varepsilon_{2i}) \right)$$

which has the conditional bias $-\frac{M}{2M-n_{12}}(\alpha_1 - \alpha_2)$, and variance $\frac{1}{2M-n_{12}} + \frac{1}{n_{12}}$. Therefore, the conditional mean squared error is given by

$$\left(-\frac{M}{2M - n_{12}} (\alpha_1 - \alpha_2) \right)^2 + \frac{1}{2M - n_{12}} + \frac{1}{n_{12}}.$$

Similar comments apply when the second cell consists of all controls.

Again, the mean squared error of interest is computed by integrating the mean squared error with respect to the distribution of (n_{11}, n_{12}) and dividing by $1 - \left(\pi^M + (1-\pi)^M\right)^2$.

B Panel Asymptotics

B.1 Covariance Matrix

We compute the joint asymptotic ($K \to \infty$) distribution of

$$\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \begin{pmatrix} \left(1\left(1 \le n_{1k} \le M-1\right) \frac{1}{n_{1k}} \sum_{i=1}^{n_{1k}} y_{1ki} - \Pr\left(1 \le n_{1k} \le M-1\right) \cdot E\left[y_{1k}\right]\right) \\ \left(1\left(1 \le n_{1k} \le M-1\right) \frac{1}{n_{0k}} \sum_{i=n_{1k}+1}^{M} y_{0ki} - \Pr\left(1 \le n_{1k} \le M-1\right) \cdot E\left[y_{0k}\right]\right) \\ \left(1\left(1 \le n_{1k} \le M-1\right) - \Pr\left(1 \le n_{1k} \le M-1\right)\right) \\ \left(1\left(1 \le n_{1k}\right) \sum_{i=1}^{n_{1k}} y_{1ki} - E\left[1\left(1 \le n_{1k}\right) n_{1k} y_{1k}\right]\right) \\ \left(1\left(n_{1k} \le M-1\right) \sum_{i=n_{1k}+1}^{M} y_{0ki} - E\left[1\left(n_{1k} \le M-1\right) n_{0k} y_{0k}\right]\right) \\ \left(1\left(1 \le n_{1k}\right) n_{1k} - E\left[1\left(1 \le n_{1k}\right) n_{1k}\right]\right) \\ \left(1\left(n_{1k} \le M-1\right) n_{0k} - E\left[1\left(n_{1k} \le M-1\right) n_{0k}\right]\right) \end{pmatrix}$$

It is easy to see that we are dealing with a sample average of zero mean i.i.d. random vectors. We therefore need to characterize the variance matrix only. Let $V(j, j')$ denote the (j, j')-th element of the variance matrix. After some tedious algebra, it can be shown that

- $V(1,1) = g(\pi, M) E\left[\sigma_{1k}^2\right] + \left(1 - \pi^M - (1-\pi)^M\right) \cdot E\left[y_{1k}^2\right] - \left(1 - \pi^M - (1-\pi)^M\right)^2 \cdot \left(E\left[y_{1k}\right]\right)^2$
- $V(2,2) = g(1-\pi, M) E\left[\sigma_{0k}^2\right] + \left(1 - \pi^M - (1-\pi)^M\right) \cdot E\left[y_{0k}^2\right] - \left(1 - \pi^M - (1-\pi)^M\right)^2 \cdot \left(E\left[y_{0k}\right]\right)^2$

- $V(3,3) = \left(1-\pi^M-(1-\pi)^M\right)\left(\pi^M+(1-\pi)^M\right)$
- $V(4,4) = M\pi E\left[\sigma_{1k}^2\right] + \left(M\pi(1-\pi)+(M\pi)^2\right)E\left[y_{1k}^2\right] - (M\pi)^2 E\left[y_{1k}\right]^2$
- $V(5,5) = M(1-\pi)E\left[\sigma_{0k}^2\right]$
  $+\left(M\pi(1-\pi)+(M(1-\pi))^2\right)E\left[y_{0k}^2\right] - (M(1-\pi))^2 E\left[y_{0k}\right]^2$
- $V(6,6) = M\pi(1-\pi)$
- $V(7,7) = M\pi(1-\pi)$
- $V(1,2) = \left(1-\pi^M-(1-\pi)^M\right)\cdot E\left[y_{1k}y_{0k}\right] - \left(1-\pi^M-(1-\pi)^M\right)^2 E\left[y_{1k}\right]E\left[y_{0k}\right]$
- $V(1,3) = \left(1-\pi^M-(1-\pi)^M\right)\left(\pi^M+(1-\pi)^M\right)E\left[y_{1k}\right]$
- $V(1,4) = \left(1-\pi^M-(1-\pi)^M\right)E\left[\sigma_{1k}^2\right] + \left(\pi M-\pi^M M\right)E\left[y_{1k}^2\right]$
  $-M\pi\left(1-\pi^M-(1-\pi)^M\right)E\left[y_{1k}\right]^2$
- $V(1,5) = \left(-(1-\pi)^M M+M-\pi M\right)E\left[y_{1k}y_{0k}\right]$
  $-\left(1-\pi^M-(1-\pi)^M\right)(M-\pi M)E\left[y_{1k}\right]E\left[y_{0k}\right]$
- $V(1,6) = \left(-\pi^M M+\pi^{M+1}M+\pi M(1-\pi)^M\right)E\left[y_{1k}\right]$
- $V(1,7) = \left(\pi^M M-\pi^{M+1}M-\pi M(1-\pi)^M\right)E\left[y_{1k}\right]$
- $V(2,3) = \left(1-\pi^M-(1-\pi)^M\right)\left(\pi^M+(1-\pi)^M\right)E\left[y_{0k}\right]$
- $V(2,4) = \left(\pi M-\pi^M\right)M\cdot E\left[y_{1k}y_{0k}\right] - \left(1-\pi^M-(1-\pi)^M\right)M\pi\cdot E\left[y_{1k}\right]E\left[y_{0k}\right]$
- $V(2,5) = \left(1-\pi^M-(1-\pi)^M\right)E\left[\sigma_{0k}^2\right] + \left(-(1-\pi)^M M+M-\pi M\right)E\left[y_{0k}^2\right]$
  $-M(1-\pi)\left(1-\pi^M-(1-\pi)^M\right)\cdot E\left[y_{0k}\right]^2$
- $V(2,6) = \left(-\pi^M M+\pi^{M+1}M+\pi M(1-\pi)^M\right)E\left[y_{0k}\right]$
- $V(2,7) = \left(\pi^M M-\pi^{M+1}M-\pi M(1-\pi)^M\right)E\left[y_{0k}\right]$
- $V(3,4) = \left(\pi^{M+1}M+\pi M(1-\pi)^M\right)E\left[y_{1k}\right]$

- $V(3,5) = \left(\pi^M M + (1-\pi)^M M - \pi^{M+1} M - \pi M (1-\pi)^M\right) E[y_{0k}]$
- $V(3,6) = \pi^{M+1} M + \pi M (1-\pi)^M$
- $V(3,7) = \pi^M M + (1-\pi)^M M - \pi^{M+1} M - \pi M (1-\pi)^M$
- $V(4,5) = \pi M (M-1)(1-\pi) E[y_{1k} y_{0k}] - M^2 \pi (1-\pi) E[y_{1k}] E[y_{0k}]$
- $V(4,6) = \pi M (1-\pi) E[y_{1k}]$
- $V(4,7) = -\pi M (1-\pi) E[y_{1k}]$
- $V(5,6) = -\pi M (1-\pi) E[y_{0k}]$
- $V(5,7) = \pi M (1-\pi) E[y_{0k}]$
- $V(6,7) = -\pi M (1-\pi)$.

B.2 Delta Method

Let

$$A = \begin{bmatrix} \frac{1}{\left(1-\pi^M-(1-\pi)^M\right)} & 0 \\ -\frac{1}{\left(1-\pi^M-(1-\pi)^M\right)} & 0 \\ -\frac{1}{\left(1-\pi^M-(1-\pi)^M\right)} \left(E[y_{1k}] - E[y_{0k}]\right) & 0 \\ 0 & \frac{1}{M\pi} \\ 0 & -\frac{1}{M(1-\pi)} \\ 0 & -\frac{1}{M\pi} E[y_{1k}] \\ 0 & \frac{1}{M(1-\pi)} E[y_{0k}] \end{bmatrix}$$

By the delta method, we have the asymptotic variance of $(b_c, b_p)'$ equal to

$$\Psi = A^T V A = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12} & \Psi_{22} \end{bmatrix}.$$

After some tedious calculation, we obtain

$$\Psi_{1,1} = \frac{g(\pi, M)}{\left(1-\pi^M-(1-\pi)^M\right)^2} E[\sigma_{1k}^2] + \frac{g(1-\pi, M)}{\left(1-\pi^M-(1-\pi)^M\right)^2} E[\sigma_{0k}^2] + \frac{\mathrm{Var}[y_{1k} - y_{0k}]}{1-\pi^M-(1-\pi)^M},$$

$$\Psi_{22} = \frac{1}{\pi M} E\left[\sigma_{1k}^2\right] + \frac{1}{M(1-\pi)} E\left[\sigma_{0k}^2\right] + \mathrm{Var}\left[y_{1k} - y_{0k}\right] + \frac{1}{M} \mathrm{Var}\left(\sqrt{\frac{1-\pi}{\pi}} y_{1k} + \sqrt{\frac{\pi}{1-\pi}} y_{0k}\right),$$

and

$$\Psi_{12} = \frac{1}{\pi M} E\left[\sigma_{1k}^2\right] + \frac{1}{M(1-\pi)} E\left[\sigma_{0k}^2\right] + \frac{1-\pi^{M-1}}{1-\pi^M-(1-\pi)^M} \mathrm{Var}\left(y_{1k}\right) + \frac{1-(1-\pi)^{M-1}}{1-\pi^M-(1-\pi)^M} \mathrm{Var}\left(y_{0k}\right) + \frac{-2+\pi^{M-1}+(1-\pi)^{M-1}}{1-\pi^M-(1-\pi)^M} \mathrm{Cov}\left(y_{1k}, y_{0k}\right).$$

References

[1] ANGRIST, J. D. (1998), "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66, pp. 249-88.

[2] ANGRIST, J. D. AND A. B. KRUEGER (1995), "Split-Sample Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business and Economic Statistics*, 13, pp.225-35.

[3] ASHENFELTER, O. A. AND D. E. CARD (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, pp. 648-60.

[4] BEKKER, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, 62, pp. 657-81.

[5] BEKKER, P. A. AND J. VAN DER PLOEG(1996), "Instrumental Variables Estimation Based on Grouped Data," University of Groningen, mimeo.

[6] CARD, D. E. AND D. SULLIVAN (1988), "Measuring the Effect of Subsidized Training on Movements In and Out of Employment," Econometrica, 56, pp. 497-530.

[7] CHAMBERLAIN, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, pp. 305 - 334.

[8] CHAMBERLAIN, G. AND Z. GRILICHES (1975), "Unobservables with a Variance-Components Structure: Ability, Schooling, and the Economic Success of Brothers," *International Economic Review* 16, pp. 422 - 449.

[9] CHAMBERLAIN, G. AND G. IMBENS (1996), "Hierarchical Bayes Models with Many Instrumental Variables," Harvard Institute of Economic Research, Discussion Paper No. 1781.

[10] CHAMBERLAIN, G. (1984), "Panel Data" in Griliches, Z. and Intriligator, M.D. eds. *Handbook of Econometrics*, Vol. II. Amsterdam: Elsevier.

[11] CHAMBERLAIN, G. (1992), "Efficiency Bounds for Semiparametric Regression", *Econometrica* 60, pp. 567 - 596.

[12] COCHRAN, W. G. (1965), "The Planning of Observational Studies of Human Populations (with Discussion)," *Journal of the Royal Statistical Society*, Series A, 128, pp. 234-266.

[13] DEATON, A. AND C. PAXSON (1998), "Economies of Scale, Household Size, and the Demand for Food," *Journal of Political Economy*, in press.

[14] DEHEJIA, R. H. AND S. WAHBA (1998), "Propensity Score Matching Methods for Non-experimental Causal Studies," NBER Working Paper No. W6729.

[15] HAHN, J. (1998), "On the Role of the Propensity Score in the Efficient Estimation of Average Treatment Effects," *Econometrica*, 66, pp. 315-332.

[16] HECKMAN, J. J., H. ICHIMURA AND P. E. TODD (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64, pp. 605-54.

[17] HECKMAN, J. J., H. ICHIMURA AND P. E. TODD (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, pp. 261 - 294.

[18] HAUSMAN, J. A., AND W. E. TAYLOR (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49, pp. 1377 - 1398.

[19] IMBENS, G. W., D. B. RUBIN, AND B. I. SACERDOTE (1997), "Estimating Income Effects: Evidence from a Survey of Lottery Players," UCLA, mimeo.

[20] MADDALA, G. S. (1971), "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica* 39, pp. 341 - 358.

[21] MUNDLAK, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica* 46, pp. 69 - 85.

[22] ROSENBAUM, P. R. (1995), *Observational studies.* New York: Springer-Verlag.

[23] ROSENBAUM, P. R. AND D. B. RUBIN (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, pp. 41-55.

[24] ROSENBAUM, P. R. AND D. B. RUBIN (1984), "Reducing Bias in Observational studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, pp. 516-524.

[25] ROSENBAUM, P. R. AND DONALD B. RUBIN (1985A), "Constructing a Control Group using Multi-variate Matching Methods that include the Propensity Score," *American Statistician*, 39:33-38.

[26] ROSENBAUM, P. R. AND DONALD B. RUBIN (1985B), "The Bias Due to Incomplete Matching," *Biometrics* 41:103-116.

[27] RUBIN, D. B. (1973), "Matching to Remove Bias in Observational Studies," *Biometrics*, 29, pp. 159-83.

[28] RUBIN, D. B. (1979), "using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, pp. 318-328.

[29] STAIGER, D. AND J. H. STOCK (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica* , 65, pp. 557-586.

[30] SWAMY, P.A.V.B. (1970), "Efficient Inference in a Random Coefficient Regression Model", *Econometrica* 38, pp. 311-323.

[31] TODD, P. E. (1995), "Local Linear Approaches to Program Evaluation Using a Semiparametric Propensity Score," University of Pennsylvania: mimeo.

[32] WALLACE, T. D., AND A. HUSSAIN (1969), "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica* 37, pp. 55 - 72.

Table 1: Actual Relative Performance (SD(b_c)/SD(b_p)) of b_p in a Two-Cell Example ($\pi = .1$)

		Covariate R^2								
		.40	.35	.30	.25	.20	.15	.10	.05	.00
	2	0.90	0.94	0.98	1.02	1.06	1.10	1.13	1.17	1.21
	3	0.84	0.88	0.92	0.95	0.99	1.02	1.06	1.09	1.13
	4	0.82	0.85	0.89	0.92	0.96	0.99	1.02	1.06	1.09
	5	0.81	0.84	0.88	0.91	0.94	0.97	1.01	1.04	1.07
	6	0.80	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.05
	7	0.80	0.83	0.86	0.90	0.93	0.96	0.99	1.02	1.05
	8	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	9	0.79	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	10	0.79	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	11	0.79	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	12	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	13	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	14	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	15	0.80	0.83	0.86	0.89	0.93	0.95	0.98	1.01	1.04
M	16	0.80	0.83	0.87	0.90	0.93	0.96	0.98	1.01	1.04
(Cell Size)	17	0.80	0.83	0.87	0.90	0.93	0.96	0.99	1.01	1.04
	18	0.80	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.04
	19	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.05
	20	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.05
	21	0.81	0.84	0.87	0.91	0.94	0.97	0.99	1.02	1.05
	22	0.81	0.84	0.88	0.91	0.94	0.97	1.00	1.02	1.05
	23	0.81	0.85	0.88	0.91	0.94	0.97	1.00	1.03	1.05
	24	0.81	0.85	0.88	0.91	0.94	0.97	1.00	1.03	1.06
	25	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.03	1.06
	26	0.82	0.85	0.88	0.91	0.95	0.97	1.00	1.03	1.06
	27	0.82	0.85	0.88	0.92	0.95	0.98	1.01	1.03	1.06
	28	0.82	0.85	0.89	0.92	0.95	0.98	1.01	1.04	1.06
	29	0.82	0.85	0.89	0.92	0.95	0.98	1.01	1.04	1.06
	30	0.82	0.86	0.89	0.92	0.95	0.98	1.01	1.04	1.07
	80	0.80	0.84	0.87	0.90	0.93	0.96	0.98	1.01	1.04
	90	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.03
	100	0.80	0.83	0.86	0.89	0.92	0.95	0.97	1.00	1.03

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model with two covariate cells. Cell size is fixed at M. The standard errors are based on an exact calculation detailed in Appendix A. The probability of treatment in this case is 1/10.

Table 2: Actual Relative Performance (SD(b_c)/SD(b_p)) of b_p in a Two-Cell Example ($\pi = .5$)

		Covariate R^2								
		.40	.35	.30	.25	.20	.15	.10	.05	.00
	2	0.91	0.95	0.98	1.02	1.05	1.08	1.11	1.14	1.17
	3	0.87	0.90	0.94	0.97	1.01	1.04	1.07	1.10	1.14
	4	0.84	0.88	0.91	0.95	0.98	1.01	1.05	1.08	1.11
	5	0.83	0.86	0.90	0.93	0.96	1.00	1.03	1.06	1.09
	6	0.82	0.85	0.89	0.92	0.95	0.98	1.02	1.05	1.08
	7	0.81	0.85	0.88	0.91	0.94	0.98	1.01	1.04	1.06
	8	0.81	0.84	0.87	0.91	0.94	0.97	1.00	1.03	1.06
	9	0.80	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.05
	10	0.80	0.83	0.86	0.90	0.93	0.96	0.99	1.01	1.04
	11	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04
	12	0.79	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.03
	13	0.79	0.82	0.86	0.89	0.92	0.95	0.97	1.00	1.03
	14	0.79	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.03
	15	0.79	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.02
M	16	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
(Cell Size)	17	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	18	0.78	0.82	0.85	0.88	0.91	0.94	0.96	0.99	1.02
	19	0.78	0.82	0.85	0.88	0.91	0.94	0.96	0.99	1.02
	20	0.78	0.82	0.85	0.88	0.91	0.93	0.96	0.99	1.02
	21	0.78	0.81	0.85	0.88	0.91	0.93	0.96	0.99	1.02
	22	0.78	0.81	0.85	0.88	0.90	0.93	0.96	0.99	1.01
	23	0.78	0.81	0.84	0.88	0.90	0.93	0.96	0.99	1.01
	24	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.01
	25	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.01
	26	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.01
	27	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.98	1.01
	28	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.98	1.01
	29	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.98	1.01
	30	0.78	0.81	0.84	0.87	0.90	0.93	0.96	0.98	1.01
	80	0.78	0.81	0.84	0.87	0.90	0.92	0.95	0.98	1.00
	90	0.78	0.81	0.84	0.87	0.90	0.92	0.95	0.98	1.00
	100	0.78	0.81	0.84	0.87	0.90	0.92	0.95	0.98	1.00

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model with two covariate cells. Cell size is fixed at M. The standard errors are based on an exact calculation detailed in Appendix A. The probability of treatment in this case is 1/2.

Table 3: Relative Performance (SD(b_c)/SD(b_p)) of b_p Using Panel-Asymptotics ($\pi = .1$)

		Covariate R^2								
		.40	.35	.30	.25	.20	.15	.10	.05	.00
	2	1.10	1.14	1.18	1.23	1.27	1.30	1.34	1.38	1.41
	3	0.95	0.99	1.03	1.06	1.10	1.13	1.16	1.19	1.23
	4	0.90	0.94	0.97	1.01	1.04	1.07	1.10	1.13	1.16
	5	0.88	0.91	0.95	0.98	1.01	1.04	1.08	1.10	1.13
	6	0.87	0.90	0.94	0.97	1.00	1.03	1.06	1.09	1.12
	7	0.86	0.90	0.93	0.96	1.00	1.03	1.06	1.09	1.11
	8	0.86	0.90	0.93	0.96	0.99	1.03	1.06	1.08	1.11
	9	0.86	0.90	0.93	0.96	0.99	1.03	1.06	1.08	1.11
	10	0.86	0.90	0.93	0.96	1.00	1.03	1.06	1.09	1.11
	11	0.86	0.90	0.93	0.97	1.00	1.03	1.06	1.09	1.12
	12	0.87	0.90	0.94	0.97	1.00	1.03	1.06	1.09	1.12
	13	0.87	0.90	0.94	0.97	1.00	1.03	1.06	1.09	1.12
	14	0.87	0.91	0.94	0.97	1.01	1.04	1.07	1.10	1.12
	15	0.87	0.91	0.94	0.98	1.01	1.04	1.07	1.10	1.13
M	16	0.88	0.91	0.95	0.98	1.01	1.04	1.07	1.10	1.13
(Cell Size)	17	0.88	0.91	0.95	0.98	1.01	1.04	1.08	1.10	1.13
	18	0.88	0.92	0.95	0.98	1.02	1.05	1.08	1.11	1.14
	19	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
	20	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
	21	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
	22	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
	23	0.89	0.92	0.96	0.99	1.02	1.06	1.09	1.12	1.14
	24	0.89	0.92	0.96	0.99	1.02	1.06	1.09	1.12	1.15
	25	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	26	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	27	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	28	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	29	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	30	0.89	0.92	0.96	0.99	1.03	1.06	1.09	1.12	1.15
	80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.04	1.07
	90	0.82	0.85	0.89	0.92	0.95	0.98	1.00	1.03	1.06
	100	0.81	0.85	0.88	0.91	0.94	0.97	1.00	1.02	1.05

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model. Cell size is fixed at M. The standard errors are based on the panel-asymptotic approximation in Theorem 1. The probability of treatment in this case is 1/10.

Table 4: Relative Performance (SD(b_c)/SD(b_p)) of b_p Using Panel-Asymptotics ($\pi = .5$)

		Covariate R^2								
		.40	.35	.30	.25	.20	.15	.10	.05	.00
	2	1.10	1.14	1.18	1.23	1.27	1.30	1.34	1.38	1.41
	3	0.95	0.99	1.03	1.06	1.10	1.13	1.16	1.19	1.23
	4	0.90	0.94	0.98	1.01	1.04	1.08	1.11	1.14	1.17
	5	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
	6	0.87	0.90	0.94	0.97	1.00	1.03	1.06	1.09	1.12
	7	0.86	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11
	8	0.85	0.88	0.91	0.95	0.98	1.01	1.04	1.07	1.09
	9	0.84	0.87	0.90	0.94	0.97	1.00	1.03	1.05	1.08
	10	0.83	0.86	0.90	0.93	0.96	0.99	1.02	1.05	1.07
	11	0.82	0.86	0.89	0.92	0.95	0.98	1.01	1.04	1.06
	12	0.82	0.85	0.88	0.92	0.95	0.97	1.00	1.03	1.06
	13	0.81	0.85	0.88	0.91	0.94	0.97	1.00	1.02	1.05
	14	0.81	0.84	0.88	0.91	0.94	0.96	0.99	1.02	1.05
	15	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.04
M	16	0.80	0.84	0.87	0.90	0.93	0.96	0.99	1.01	1.04
(Cell Size)	17	0.80	0.84	0.87	0.90	0.93	0.96	0.98	1.01	1.04
	18	0.80	0.83	0.86	0.90	0.92	0.95	0.98	1.01	1.03
	19	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.03
	20	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.00	1.03
	21	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.00	1.03
	22	0.80	0.83	0.86	0.89	0.92	0.95	0.97	1.00	1.03
	23	0.79	0.83	0.86	0.89	0.92	0.95	0.97	1.00	1.03
	24	0.79	0.83	0.86	0.89	0.92	0.94	0.97	1.00	1.02
	25	0.79	0.82	0.86	0.89	0.91	0.94	0.97	1.00	1.02
	26	0.79	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.02
	27	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	28	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	29	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	30	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	80	0.78	0.81	0.84	0.87	0.90	0.93	0.95	0.98	1.01
	90	0.78	0.81	0.84	0.87	0.90	0.93	0.95	0.98	1.01
	100	0.78	0.81	0.84	0.87	0.90	0.93	0.95	0.98	1.01

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model. Cell size is fixed at M. The standard errors are based on the panel-asymptotic approximation in Theorem 1. The probability of treatment in this case is 1/2.

Table 5: Monte Carlo Relative Performance (SD(b_c)/SD(b_p)) of b_p ($\pi = .1$)

	Covariate R^2								
	.40	.35	.30	.25	.20	.15	.10	.05	.00
2	1.06	1.10	1.14	1.18	1.22	1.26	1.29	1.33	1.36
3	0.97	1.01	1.04	1.08	1.12	1.15	1.18	1.22	1.25
4	0.93	0.96	1.00	1.04	1.07	1.10	1.14	1.17	1.20
5	0.90	0.94	0.98	1.01	1.04	1.08	1.11	1.14	1.17
6	0.89	0.93	0.97	1.00	1.03	1.06	1.09	1.12	1.15
7	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
8	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
9	0.88	0.92	0.95	0.98	1.02	1.05	1.08	1.11	1.13
10	0.88	0.92	0.95	0.98	1.01	1.04	1.07	1.10	1.13
11	0.88	0.91	0.95	0.98	1.01	1.05	1.08	1.10	1.13
12	0.88	0.91	0.95	0.98	1.01	1.04	1.07	1.10	1.13
13	0.88	0.91	0.95	0.98	1.01	1.05	1.08	1.10	1.13
14	0.88	0.91	0.95	0.98	1.01	1.05	1.08	1.10	1.13
15	0.88	0.92	0.95	0.98	1.02	1.05	1.08	1.11	1.13
M^* 16	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
Avg. Cell Size 17	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
18	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.14
19	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
20	0.88	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
21	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
22	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
23	0.89	0.92	0.96	0.99	1.02	1.06	1.09	1.11	1.14
24	0.88	0.92	0.96	0.99	1.02	1.05	1.09	1.12	1.15
25	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
26	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
27	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
28	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
29	0.89	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14
30	0.88	0.92	0.96	0.99	1.02	1.05	1.08	1.11	1.14

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model. Cell size is random. The standard errors were calculated by Monte-Carlo integration of analytic formulas that condition on cell sizes and number treated. The probability of treatment in this case is 1/10.

Table 6: Monte Carlo Relative Performance (SD(b_c)/SD(b_p)) of b_p ($\pi = .5$)

		Covariate R^2								
		.40	.35	.30	.25	.20	.15	.10	.05	.00
	2	1.04	1.08	1.12	1.16	1.20	1.24	1.27	1.31	1.34
	3	0.95	0.99	1.03	1.06	1.10	1.13	1.16	1.19	1.22
	4	0.91	0.94	0.98	1.01	1.05	1.08	1.11	1.14	1.17
	5	0.88	0.92	0.95	0.99	1.02	1.05	1.08	1.11	1.13
	6	0.87	0.90	0.93	0.97	1.00	1.03	1.06	1.09	1.12
	7	0.85	0.89	0.92	0.95	0.98	1.01	1.04	1.07	1.10
	8	0.84	0.88	0.91	0.94	0.97	1.00	1.03	1.06	1.09
	9	0.84	0.87	0.90	0.93	0.96	0.99	1.02	1.05	1.08
	10	0.83	0.86	0.90	0.93	0.96	0.99	1.02	1.04	1.07
	11	0.82	0.86	0.89	0.92	0.95	0.98	1.01	1.04	1.06
	12	0.82	0.85	0.88	0.92	0.95	0.97	1.00	1.03	1.06
	13	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.02	1.05
	14	0.81	0.85	0.88	0.91	0.94	0.97	0.99	1.02	1.05
	15	0.81	0.84	0.88	0.91	0.93	0.96	0.99	1.02	1.04
M^*	16	0.81	0.84	0.87	0.90	0.93	0.96	0.99	1.01	1.04
Avg. Cell Size	17	0.80	0.84	0.87	0.90	0.93	0.96	0.98	1.01	1.04
	18	0.80	0.84	0.87	0.90	0.93	0.95	0.98	1.01	1.03
	19	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.01	1.03
	20	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.00	1.03
	21	0.80	0.83	0.86	0.89	0.92	0.95	0.98	1.00	1.03
	22	0.80	0.83	0.86	0.89	0.92	0.95	0.97	1.00	1.03
	23	0.80	0.83	0.86	0.89	0.92	0.95	0.97	1.00	1.03
	24	0.79	0.83	0.86	0.89	0.92	0.94	0.97	1.00	1.02
	25	0.79	0.83	0.86	0.89	0.92	0.94	0.97	1.00	1.02
	26	0.79	0.83	0.86	0.89	0.91	0.94	0.97	1.00	1.02
	27	0.79	0.82	0.85	0.88	0.91	0.94	0.97	1.00	1.02
	28	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	29	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02
	30	0.79	0.82	0.85	0.88	0.91	0.94	0.97	0.99	1.02

Notes: The table reports relative standard errors for covariate-matching and propensity-score estimates of a constant-treatment-effect homoscedastic model. Cell size is random. The standard errors were calculated by Monte-Carlo integration of analytic formulas that condition on cell sizes and number treated. The probability of treatment in this case is 1/2.

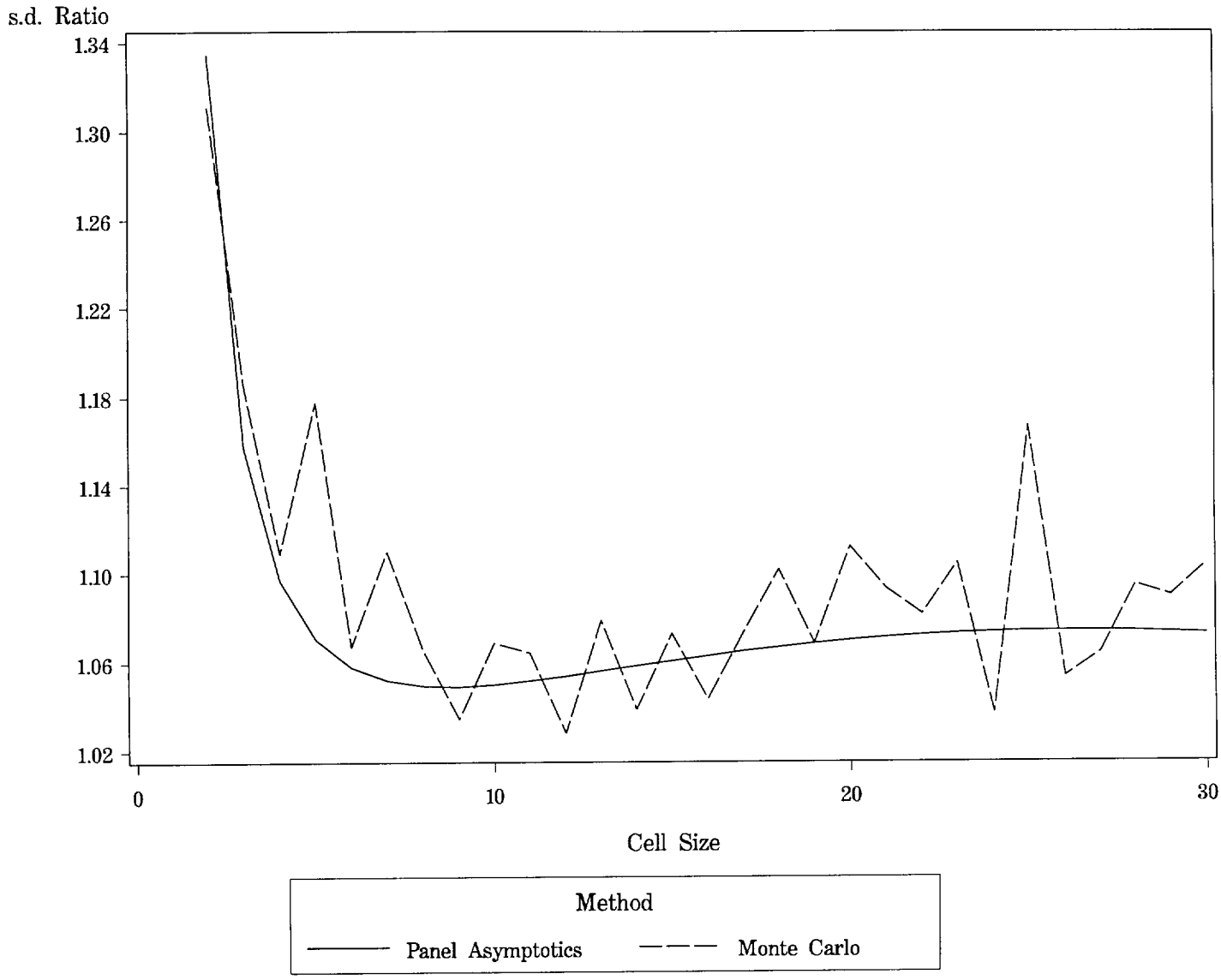

Figure 1. Panel−asymptotic vs. Monte Carlo s.d.(Bc)/s.d.(Bp) for expected cell sizes 2−30
Monte Carlo design: 500 replications, 30 cells;
Covariate Rsquare for y0=.1, treatment probability=.1
Heterogeneous treatment effect equal to 0 or 1 with probability .5

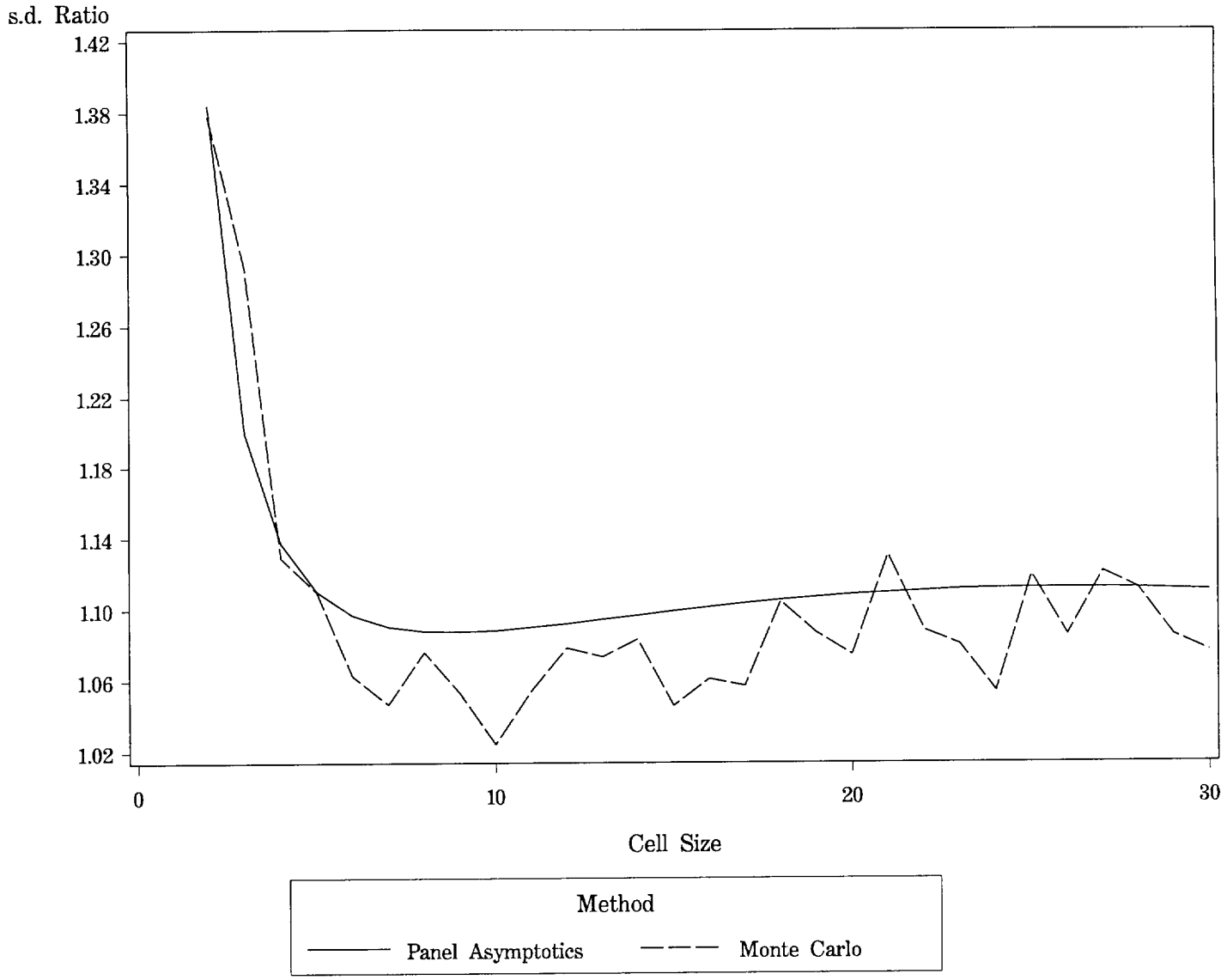

Figure 2. Panel−asymptotic vs. Monte Carlo s.d.(Bc)/s.d.(Bp) for expected cell sizes 2−30
Monte Carlo design: 500 replications, 30 cells;
Covariate Rsquare for y0=.1, treatment probability=.1
Treatment effect equals 0 or 1, negatively correlated with y0