IFAU
Institute for Evaluation of Labour Market and Education Policy

Effects of correlated covariates on the efficiency of matching and inverse probability weighting estimators for causal inference

Ronnie Pingel
Ingeborg Waernbaum

WORKING PAPER 2013:5

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166

Effects of correlated covariates on the efficiency of matching and inverse probability weighting estimators for causal inference*

Ronnie Pingel[a] and Ingeborg Waernbaum[b]

[a]Department of Statistics, Uppsala University, Uppsala, Sweden,
[b]Department of Statistics, Umeå University, Umeå, Sweden,
and
Institute for Evaluation of
Labour Market and Education Policy, Uppsala, Sweden

February 13, 2013

Summary

In observational studies the overall aim when fitting a model for the propensity score is to reduce bias for an estimator of the causal effect. For this purpose guidelines for covariate selection for propensity score models have been proposed in the causal inference literature. To make the assumption of an unconfounded treatment plausible researchers might be tempted to include many, possibly correlated, covariates in the propensity score model. In this paper we study how the efficiency of matching and inverse probability weighting estimators for average causal effects change when the covariates are correlated. We investigate the case with multivariate normal covariates and linear models for the propensity score and potential outcomes and show results under different model assumptions. We show that the correlation can both increase and decrease the large sample variances of the estimators, and that the correlation affects the efficiency of the estimators differently, both with regard to direction and magnitude. Moreover, the strength of the confounding towards the outcome and the treatment plays an important role.

Keywords: Efficiency bound, observational study, propensity score, variable selection
JEL-codes: C13; C40; C52

*We thank conference participants at the 26th International Biometric Conference in Kobe and seminar participants at the joint IFAU and Department of Statistics seminar. We would also like to thank Marcus Frolich and Per Johansson for valuable comments.

Contents

1 Introduction

Drawing inference about the causal effect of a treatment with observational data is a common goal in empirical research. An assumption of unconfoundedness is fulfilled if the treatment assignment is independent of the potential outcomes resulting from assignment to treatment or no treatment when conditioning on a set of observed pre-treatment variables. Estimators adjusting for confounding variables are used to disentangle the treatment effect from the effect of other variables, in the following referred to as covariates, that systematically differ between treated and controls. Instead of controlling for the covariates directly it is sufficient to control for the propensity score, the conditional probability of treatment given the covariates (Rosenbaum and Rubin 1983). Propensity score methods have gained wide applicability in both the social and medical sciences, see e.g., Thoemmes and Kim (2011) and Luo, Gardiner, and Bradly (2010) for recent reviews.

For the data analyst a central component in the estimation process is to specify the propensity score model. The topic has been addressed in the literature with focus on both variable selection (Heckman, Ichimura, and Todd 1998; Lunceford and Davidian 2004; Brookhart, Schneeweiss, Rothman, Glynn, Avorn, and Stürmer 2006) and functional form specification (Dehejia and Wahba 1999; Millimet and Tchernis 2009; Waernbaum 2010; Waernbaum 2012). Through simulations Millimet and Tchernis (2009) study the case of overfitting in terms of including both irrelevant variables and higher order terms as well as underfitting in terms of excluding relevant covariates and higher order terms. Their conclusions were in general positive towards overfitting and they found that the cost in terms of efficiency was small in the simulations under study.

In this paper we address the setting where the researcher has a rich set of data available with many covariates describing similar characteristics of the individual, e.g., there may be several variables describing health or socio-economic conditions of the individiual. In order to justify the unconfoundedness assumption it could be considered appealing to include as many covariates as possible. In practice, the analyst could choose one or several variables describing for instance the individuals' education. In such cases the covariates will most likely be strongly correlated. For matching estimators it has been suggested that collinearity in a propensity score model is not of concern since inference regarding the individual parameters in the model is not of primary interest and the collinearity does not per se reduce the predicitive power of the propensity score (Stuart 2010).

The results in this paper contribute to the discussion of overfitting by studying how the efficiency is affected by using a propensity score model with variables that are correlated. For this purpose we study two different approaches of estimation of causal effects with propensity scores. Estimators such as matching or stratification estimators use the property of the propensity score as a balancing score, i.e., within the matched sets or strata the covariates of the treated and controls follow the same distribution (see Imbens and Wooldridge 2009 for a review of matching and stratification estimators). A different approach is to use the propensity scores to generate the missing potential outcomes as in inverse probability weighting (IPW) (Wooldridge 2007). Here, we study and compare the large sample variances of the matching estimator as described in Abadie and Imbens (2006) and an IPW estimator described in Hirano, Imbens, and Ridder (2003) and Lunceford and Davidian (2004). To investigate the impact of the correlation as such on the asymptotic variance of the estimators we assume parametric models for the data generating process (DGP), also we assume that the covariates are normally distributed.

In our setting we find that the variance of the matching estimator and the IPW estimator can be either larger or smaller when the covariates are correlated depending on the regression coefficients in the outcome and propensity score models. Also, we see that the impact of the correlation can be in different directions, i.e., for the same values of the parameters in the models a change in the correlation can have the opposite effect on the IPW estimator compared to the matching estimator. For instance, if the regression coefficients are either all positive or all negative the variance of the IPW estimator increases if the correlation increases which is not necessarily the case for the matching estimator under study. We study the relative efficiency for the estimators and show results on the maximum and minimum values of the ratio of the large sample variances together with results of the rate of change of the ratio with respect to the correlation. Under the conditions studied our results indicate that the asymptotic variance of the matching estimator is less affected by an increase in the correlation than the IPW estimator.

We proceed as follows: In Section 2, we present the theoretical framework, notation and the matching and IPW estimators. Section 3 presents the results of the paper concerning correlation and efficiency of the previously described estimators and Section 4 concludes with a discussion.

2 Model and theory

In this section, we describe the Neyman-Rubin framework of causal inference (Neyman 1923; Rubin 1974) through potential outcomes. For a given population we assume that we have a random sample of N individuals $i = 1, \ldots, N$. We consider a binary treatment, $W = 1$ if the individual is treated and $W = 0$ if the individual is a control. The definition of a causal effect of the treatment on a response variable of interest is the difference of the potential outcomes $Y_1 - Y_0$, where Y_1 is the potential outcome under treatment and Y_0 is the potential outcome under the control treatment. In this paper interest lies in estimation of the average causal effect, $\tau = E(Y_1 - Y_0)$, under an assumption of unconfoundedness. We denote by X a vector of pre-treatment variables referred to as covariates in the following. To summarize, for each individual we observe the variables (X, W, Y) where Y is the observed response variable $Y = WY_1 + (1 - W)Y_0$. When the treatment is not randomized the average causal effect can be identified under an assumption of strong ignorability (Rosenbaum and Rubin 1983),

Assumption 1 (Strong ignorability)
(i) (unconfoundedness) $(Y_1, Y_0) \perp\!\!\!\perp W \mid X$
(ii) (overlap) $0 < P(W = 1|X) < 1$.

Throughout the paper we assume that Assumption 1 holds although milder versions are sufficient for the identification of an average causal effect, see e.g., Wooldridge (2002, Chapter 18) for alternative versions of identifying assumptions. Under Assumption 1 the population average causal effect can be identified with the observed data by comparing treated and controls conditional on X and then take the average of the conditional differences.

$$\tau = E(Y_1 - Y_0) = E(E(Y_1 \mid X, W = 1) - E(Y_0 \mid X, W = 0)). \quad (1)$$

Instead of comparing treated and controls with the same values on all covariates it suffices to condition on the propensity score, $p(X) \equiv P(W = 1 \mid X)$ (Rosenbaum and Rubin 1983), a scalar function of the covariates, hence overcoming the curse of dimensionality present in all nonparametric estimators. As an alternative the propensity score can be used as weights to generate the missing potential outcomes

$$\tau = E(Y_1 - Y_0) = E\left(\frac{WY}{p(X)}\right) - E\left(\frac{(1 - W)Y}{1 - p(X)}\right). \quad (2)$$

Here, the propensity score weights the observed data to correct for the disproportionality of the observed responses with respect to the potential outcomes in the population similar to sample weighting proposed by Horvitz and Thompson (1952). The two representations displayed in (1) and (2) differ with respect to how the propensity score is used. When conditioning as in Equation 1 the propensity score or any other balancing score, $b(X)$ such that $X \perp\!\!\!\perp W \mid b(X)$, can be used instead of the full covariate vector X whereas in Equation 2 the propensity score is used to generate the $1/p(X) - 1$ missing potential outcomes with similar characteristics. Throughout this paper we assume that the propensity score is a known function although in practice it is often estimated.

Irrespective of whether the propensity score is known or estimated, Hahn (1998) shows that the semi-parametric asymptotic efficiency bound for the average treatment effect is

$$\sigma^2_{\text{EFF}} = E\left[(E(Y_1 - Y_0|X) - \tau)^2 + \frac{V(Y_1|X)}{p(X)} + \frac{V(Y_0|X)}{1-p(X)}\right]. \tag{3}$$

Efficient estimators, i.e., semi-parametric estimators reaching (3) have been proposed, see e.g., Hahn (1998), Hirano, Imbens, and Ridder (2003), Imbens, Newey, and Ridder (2005).

2.1 Inverse probability weighting estimators

The estimation approach illustrated by Equation 2 has formed the basis for IPW estimators. The IPW estimators belong to a class of efficient semiparametric estimators in the missing data framework, see e.g., Tsiatis (2006). A straightforward estimator would be to consider

$$\frac{1}{N}\sum_{i=1}^{N}\frac{W_iY_i}{p(X_i)} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-W_i)Y_i}{1-p(X_i)}.$$

However, since the weights do not always add up to one this estimator is not particularly appealing, see e.g., Hirano, Imbens, and Ridder (2003) for further discussion. Instead we study the following estimator using normalized weights proposed by Hirano, Imbens, and Ridder (2003) and also described in Lunceford and Davidian (2004):

$$\hat{\tau}_{\text{IPW}} = \left(\sum_{i=1}^{N}\frac{W_i}{p(X_i)}\right)^{-1}\sum_{i=1}^{N}\frac{W_iY_i}{p(X_i)} - \left(\sum_{i=1}^{N}\frac{1-W_i}{1-p(X_i)}\right)^{-1}\sum_{i=1}^{N}\frac{(1-W_i)Y_i}{1-p(X_i)}. \tag{4}$$

Its large sample distribution is

$$\sqrt{N}\left(\hat{\tau}_{\text{IPW}} - \tau\right) \xrightarrow{d} N\left(0, \sigma^2_{\text{IPW}}\right)$$

with variance

$$\sigma^2_{\text{IPW}} = E\left[\frac{(Y_1-\mu_1)^2}{p(X)} + \frac{(Y_0-\mu_0)^2}{1-p(X)}\right], \tag{5}$$

where $\mu_w = E(Y_w)$, $w = 0, 1$.

IPW estimators are easily implemented in practice, but a concern is the risk of unstable estimates when $p(X)$ is close to zero or one, see e.g. the discussion in Kang and Schafer (2007). It should be noted that neither the unnormalized estimator nor the normalized version in (4) achieve the semi-parametric efficiency bound when using the true propensity score.

2.2 The matching estimator

A matching estimator imputes the missing potential outcome from one or several nearest neighbours from the opposite treatment group. Below, we study a matching estimator that compares outcomes between treated and controls, with each observation matched to a fixed number of units with the opposite treatment. The matching is done with replacement, allowing each unit to be used as a match multiple times. The sample consists of N_1 treated and N_0 control units with $N_1 + N_0 = N$. Assuming that at least one covariate and the function $p(\cdot)$ being continuous, the propensity score is also continuous with no ties present. For individuals i and i' from opposite treatments we denote the distance $d_{ii'}$,

$$d_{ii'} = \mid p(X_i) - p(X_{i'}) \mid . \tag{6}$$

where for each i we denote by J_i a set $J_i = \{1, 2, \ldots, i', \ldots, M\}$ of indices of the M individuals with the smallest order statistics $d_{i(i')}$, $i' \leq M$. The matching estimator, $\hat{\tau}_{\text{M}}$, matching treated and controls to a fixed number of M matches is defined

$$\hat{\tau}_{\text{M}} = \frac{1}{N}\sum_{i=1}^{N} W_i(Y_i - \hat{Y}_{0i}) + (1-W_i)(\hat{Y}_{1i} - Y_i), \tag{7}$$

where $\hat{Y}_{0i} = \frac{1}{M}\sum_{i' \in J_i} Y_{i'}$ and $\hat{Y}_{1i} = \frac{1}{M}\sum_{i' \in J_i} Y_{i'}$ are the observed response means for the M individuals with the smallest absolute difference in the propensity score (6).

Abadie and Imbens (2006) show that the large sample distribution of the matching estimator based on the true propensity score is[1]

$$\sqrt{N}\left(\hat{\tau}_{\text{M}} - \tau\right) \xrightarrow{d} N\left(0, \sigma^2_{\text{M}}\right)$$

[1]Besides Assumption 1 further assumptions regarding differentiability and support of the propensity score are needed. See Abadie and Imbens (2006) for details.

with variance

$$\sigma^2_{\text{M}} = E\left[\left(E\left[Y|W=1,p(X)\right] - E\left[Y|W=0,p(X)\right] - \tau\right)^2\right]$$
$$+ E\left[V\left[Y|W=1,p(X)\right]\left(\frac{1}{p(X)} + \frac{1}{2M}\left(\frac{1}{p(X)} - p(X)\right)\right)\right]$$
$$+ E\left[V\left[Y|W=0,p(X)\right]\left(\frac{1}{1-p(X)} + \frac{1}{2M}\left(\frac{1}{1-p(X)} - (1-p(X))\right)\right)\right]. \tag{8}$$

Matching estimators are intuitive to use for practitioners with a smoothing parameter, i.e., the number of matches, that is easy to interpret. Although matching estimators are biased in general, where using a single match gives the least bias, the bias is of sufficiently low enough order to be ignored when matching on a scalar, such as the propensity score. The large sample variance of the matching estimator and the asymptotic variance bound when matching on the propensity score are then related as follows

$$\frac{\sigma^2_{\text{M}} - \sigma^2_{\text{EFF}}}{\sigma^2_{\text{EFF}}} < \frac{1}{2M}.$$

Obviously, σ^2_{M} does not achieve the variance bound, but increasing the number of matches improves the efficiency of the estimator.

3 Correlated covariates and efficiency

In this section we investigate the behaviour of the estimators defined in (4) and (7) when the covariates are correlated. The estimators are also compared to the efficiency bound in (3). For this purpose we assume parametric models for the data generating process. The analysis of the estimators that follows is then divided into two parts. The first part focuses on correlation and the behaviour of the large sample variances σ^2_{IPW} and σ^2_{M}, while the second part studies how correlation affects the relative efficiency of the estimators, $R = \sigma^2_{\text{M}}/\sigma^2_{\text{IPW}}$.

3.1 The data generating process

To investigate the large sample variances of the matching and IPW estimator we assume linear models generating the potential outcomes and a logistic model generating the propensity score.

Assumption 2 (Outcome model) *Each potential outcome, w, is generated by a linear model*

$$Y_w = \alpha_w + \beta'_w X + \varepsilon_w, \qquad w = 0, 1, \tag{9}$$

with an error term, ε_w, a k-dimensional parameter vector β_w with elements β_{wj} for $j = 1, \ldots, k$, and a k-dimensional random variable, X.

Assumption 3 (Treatment assignment) *The propensity score is generated by a logistic model*

$$p(\gamma' X) = \frac{e^{\gamma' X}}{1 + e^{\gamma' X}}, \tag{10}$$

where $\gamma = (\gamma_1, \ldots, \gamma_j, \ldots, \gamma_k)$ is a coefficent vector.

Further, we also make an assumption on the distribution of the covariates.

Assumption 4 (Distribution of the covariates) *The covariate vector, X, has a k-variate normal distribution with mean vector θ and covariance matrix Σ of full rank. Furthermore, the error term ε_w is uncorrelated with X and follows a normal distribution with mean zero and variance $\sigma^2_{\varepsilon_w}$.*

By the assumption of multivariate normality, $\gamma' X = Z$ has a normal distribution with mean $\gamma'\theta = \mu_Z$ and variance $\gamma'\Sigma\gamma = \sigma^2_Z$. Similarly, Y_w has a normal distribution with mean $\mu_{Y_w} = \alpha_w + \beta'_w\theta$ and variance $\sigma^2_{Y_w} = \beta'_w\Sigma\beta_w + \sigma^2_{\varepsilon_w}$. The covariance between Y_w and Z is $\sigma_{Y_w,Z} = \beta'\Sigma\gamma$. In this setting, W is a Bernoulli random variable with success probability $p(Z)$. Finally, note that the average treatment effect is given by

$$\tau = \alpha_1 - \alpha_0 + \theta'(\beta_1 - \beta_0). \tag{11}$$

3.2 Correlation and the behavior of the estimators

We now turn to the analysis of how correlation among the covariates affects the variances of the estimators defined in (4) and (7). We begin by establishing the following proposition regarding the large sample variance of the inverse probability weighting estimator in (5).

Proposition 1 *Under Assumptions $1 - 4$ the large sample variance of $\hat{\tau}_{IPW}$ is*

$$\sigma^2_{IPW} = \sigma^2_{Y_1} + e^{-\mu_Z + \sigma^2_Z/2}\left(\sigma^2_{Y_1} + \sigma^2_{Y_1,Z}\right) + \sigma^2_{Y_0} + e^{\mu_Z + \sigma^2_Z/2}\left(\sigma^2_{Y_0} + \sigma^2_{Y_0,Z}\right). \tag{12}$$

The proofs to this and all of the following propositions and corollaries are provided in the Appendix. The parts in (12) involving Σ are $\sigma^2_{Y_w}$, σ^2_Z and $\sigma^2_{Y_w,Z}$. Recalling the data generating process, we conclude by inspection that as the correlation in a set of variables increases and if all elements in β and γ share the same sign, then σ^2_{IPW} also increases. For the case when some elements in β and γ are strictly positive and some elements are strictly negative, σ^2_{IPW} may increase or decrease as the correlation among the variables increases.

Next, we state a proposition regarding the variance of the matching estimator.

Proposition 2 *Under Assumptions* $1-4$ *the large sample variance of* $\hat{\tau}_M$ *is*

$$\begin{aligned}\sigma^2_M =& \frac{\sigma^2_{Y_1,Z} + \sigma^2_{Y_0,Z} - 2\sigma_{Y_1,Z}\sigma_{Y_0,Z}}{\sigma^2_Z} \\ &+ \left(\sigma^2_{Y_1} - \frac{\sigma^2_{Y_1,Z}}{\sigma^2_Z}\right) E\left[1 + \frac{e^{-Z}(1+2M)}{2M} + \frac{1}{2M(1+e^Z)}\right] \\ &+ \left(\sigma^2_{Y_0} - \frac{\sigma^2_{Y_0,Z}}{\sigma^2_Z}\right) E\left[\frac{2M+1}{2M} + \frac{e^Z(1+2M)}{2M} - \frac{1}{2M(1+e^Z)}\right]. \end{aligned} \quad (13)$$

There is no closed form available for the expectations in (13), therefore some simplifying assumptions are introduced to facilitate further analysis.

Assumption 5
(i) Constant treatment effect, i.e. $\beta_1 = \beta_0$*, and identically distributed error terms, implying that* $\sigma^2_{Y_1} = \sigma^2_{Y_0}$ *and* $\sigma_{Y_1,Z} = \sigma_{Y_0,Z}$*.*
(ii) Z *has mean zero, i.e.* $\mu_Z = 0$*.*
(iii) Σ *is an equicorrelation matrix with* $\rho \geq 0$*.*

Before using the above assumptions to establish further results regarding the behaviour σ^2_{M}, note that under Assumptions $1-5(i)$ the efficiency bound in (3) is

$$\sigma^2_{\text{EFF}} = \sigma^2_{\varepsilon_w}\left(2 + e^{-\mu_Z + \frac{1}{2}\sigma^2_Z} + e^{\mu_Z + \frac{1}{2}\sigma^2_Z}\right). \quad (14)$$

The behaviour of the efficiency bound is only affected by the treatment assignment and if all elements in γ share the same sign, then σ^2_{EFF} also increases as the correlation increases.

We now establish the following corollary regarding σ^2_{M} and the derivative of σ^2_{M} with respect to ρ.

Corollary 1 *Suppose that Assumptions* $1-4$ *hold.*
(i) Adding Assumption $5(i-ii)$*, the large sample variance of* $\hat{\tau}_M$ *is*

$$\sigma^2_M = \left(\sigma^2_{Y_w} - \frac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)\left(\frac{1 + 4M + (2+4M)e^{\frac{1}{2}\sigma^2_Z}}{2M}\right), \quad w = 0,1. \quad (15)$$

(ii) Adding Assumption $5(i-iii)$, the derivative of $\sigma^2_{\scriptscriptstyle M}$ with respect to ρ is

$$\frac{\partial\sigma^2_{\scriptscriptstyle M}}{\partial\rho} = \frac{\left(1+4M+(2+4M)e^{\frac{1}{2}\sigma^2_Z}\right)\left(\beta'_w A\beta_w + \frac{\sigma^2_{Y_w,Z}\gamma' A\gamma}{(\sigma^2_Z)^2} - \frac{2\beta'_w A\gamma\sigma_{Y_w,Z}}{\sigma^2_Z}\right)}{2M} + \left(\sigma^2_{Y_w} - \frac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)\frac{(2+4M)\gamma' A\gamma e^{\frac{1}{2}\sigma^2_Z}}{4M} \tag{16}$$

where A is a $k\times k$ hollow matrix with all diagonal elements equal to zero and all other elements equal to one.

Observe that due to Assumption $5(i)$ it is not necessary to specify which potential outcome we are referring to. Even though further assumptions have been made, it is still difficult to generalize how a change in the correlation between the covariates will affect σ^2_{M}. Therefore, we proceed with studying some examples under Assumptions $1-5(i-iii)$.

Example 1 *Let all elements in γ be equal. Then (16) can be written as the linear function*

$$\frac{\partial\sigma^2_{\scriptscriptstyle M}}{\partial\rho} = f(M,k,\gamma_j,\rho) + g(M,k,\gamma_j,\rho)\sum_{j=1}^{k-1}\sum_{j'=j+1}^{k}(\beta_{wj}-\beta_{wj'})^2 \tag{17}$$

where $f(\cdot)$ has the same sign as γ_j and $g(\cdot)$ can be positive or negative. If $f(\cdot)$ and $g(\cdot)$ have opposite signs the magnitude of the differences between elements in β_w will determine the sign of the derivative. Three numerical examples provides an illustration:

$$f(1,4,\sqrt{1/3},0) = 11.69, \quad g(1,4,\sqrt{1/3},0) = 0.84,$$
$$f(1,4,\sqrt{1/3},0.9) = 70.70, \quad g(1,4,\sqrt{1/3},0.9) = -7.70,$$
$$f(1,4,\sqrt{1/20},0) = 0.99, \quad g(1,4,\sqrt{1/20},0) = -1.21.$$

In the case of all elements in β_w being equal, then $\sum_{j=1}^{k-1}\sum_{j'=j+1}^{k}(\beta_{wj}-\beta_{wj'})^2 = 0$. An increase in the correlation will then always lead to an increase in $\sigma^2_{\scriptscriptstyle M}$ if γ_j is positive. The effect is the opposite if γ_j is negative.

Example 1 highlights two features of the behavior of the matching estimator. To begin with, it demonstrates that the variance may actually decrease as the correlation in a set of variables increases even if all elements in β_w and γ are strictly positive. This contrasts the behavior of the inverse probability weighting estimator. Furthermore, it shows the importance of different parameter values in β_w when all elements in γ are equal.

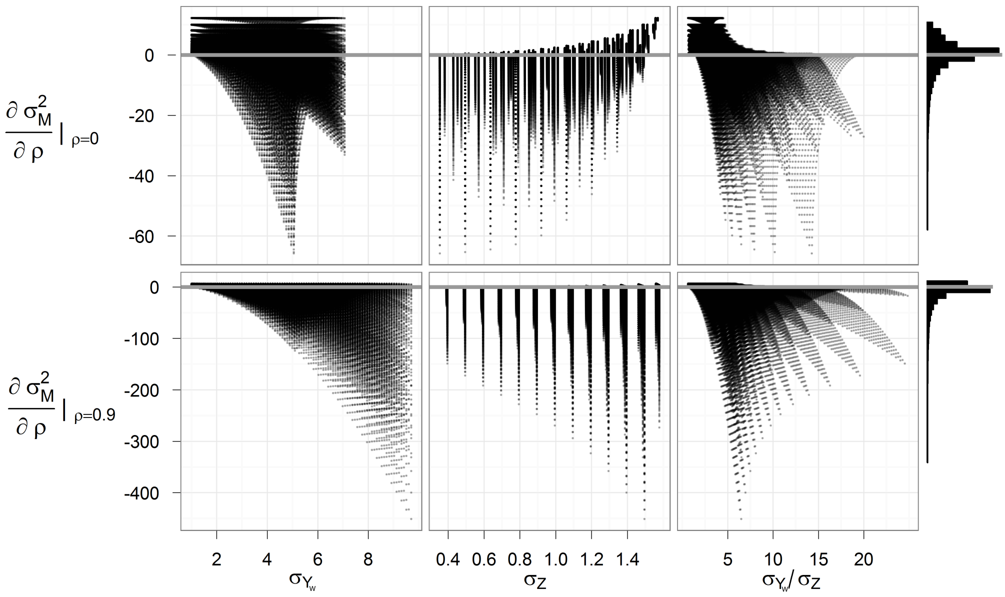


Figure 1: The relationship between σ_{Y_w}, σ_Z and σ_{Y_w}/σ_Z and $\partial\sigma^2_{\text{M}}/\partial\rho$ evaluated at $\rho = 0$ and $\rho = 0.9$. Histograms of the evaluated derivatives are displayed to the farthest right.

In the following example we study σ^2_{M} for a wider range of parameter values.

Example 2 *Let $\tau = 3$ and $\sigma^2_{\varepsilon_w} = 1$. Also, let $M = 1$ and $k = 2$. Thus the components $\beta_{w1}, \beta_{w2}, \gamma_1, \gamma_2$ and ρ influence the derivative in (16). Next, let β_{w1} and β_{w2} take the 50 values, $\{0.05, 0.15, ..., 4.95\}$, γ_1 and γ_2 take the 15 values, $\{0.05, 0.15, ..., 1.45\}$, and ρ take the values, $\{0, 0.9\}$. Select all the unique combinations and then, to avoid extreme distributions of the propensity score (see the Appendix), delete observations with $\sigma^2_Z < 0.1$ and $\sigma^2_Z > 2.5$. Use the remaining 122,400 data points for $\rho = 0$ and 84,150 data points for $\rho = 0.9$ and analyse how σ_{Y_w}, σ_Z and σ_{Y_w}/σ_Z are associated with (16). Due to scale issues the standard deviations are used instead of the variances.*

The left panels in Figure 1 show that it seems more likely for the derivative to be negative if σ_{Y_w} is large. The results in the middle panels are ambiguous. The association between σ_Z and the derivative is unclear and appears to depend on the level of ρ, where for $\rho = 0.9$ most values *are negative. The right panels display that it is more likely for the derivative to be negative if σ_{Y_w}/σ_Z is large. Furthermore, the vertical histograms to the right show that larger ρ makes it more likely for the derivative to be negative. This means that the variance of the matching estimator will decrease if the correlation increases.*

One way to interpret the results in Example 2 is to view σ_{Y_w} and σ_Z as proxies for the size of the parameters, i.e., how strongly related the covariates are to treatment assignment and outcome. Given that the parameters are positive, the example points out three important things. First, the two left panels suggest that an increase in the correlation between the covariates is more likely to decrease the variance of the matching estimator if the covariates are strongly related to the outcome. Second, by looking at the panels to the right it seems that if the covariates are strongly related to the outcome relative to how strongly related the covariates are to the treatment assignment, then increasing the correlation is likely to decrease the variance of the matching estimator. Third, by simply comparing the two vertical histograms it looks like an increase in the correlation will make the covariates more strongly related to both the treatment assignment and the outcome. Hence, a stronger initial level of correlation then makes it more likely to decrease the variance of the matching estimator when the correlation increases.

Next we consider an example that provides a more detailed view of how correlation between covariates affects the variance of the estimators. Here we separate the components in the correlation ρ, the components in the outcome β_w and the components in the treatment assignment γ. The variances are also compared to the efficiency bound in (3).

Example 3 *Assume that Assumptions $1-5(i-iii)$ hold and let $\tau = 3$, $\sigma^2_{\varepsilon_w} = 1$, $M = 1$ and $k = 2$. Evaluate (12), (15) and (3) for different values of ρ, β_{w1}, β_{w2}, γ_1 and γ_2 and plot the association between ρ and the variances and the efficiency bound.*

In Figure 2, we once again observe that the matching estimator and the IPW estimator may display opposite behaviour with respect to the effect of correlation. Also seen, and consistent with previous results, is that in the case when all elements in γ are equal and all elements in β_w are equal, the derivative of σ^2_M with respect to ρ is positive. In addition, Figure 2 provides new findings. The figure suggests that the matching estimator is less sensitive to a change in the correlation compared to the IPW estimator. Furthermore, the largest decrease in the variance of the matching estimator occurs when γ_1 is small and β_{w1} is large, while at the same time γ_2 is *large and β_{w2} is small. Interestingly, the variance of the matching estimators seems for many of these cases actually approach the efficiency bound when ρ approaches one. Moreover, Figure 2 shows that even though the IPW estimator and the efficiency bound exhibit similar behaviour, the IPW estimator is always more sensitive to a change in the correlation and the efficiency bound is only affected by ρ and γ.*

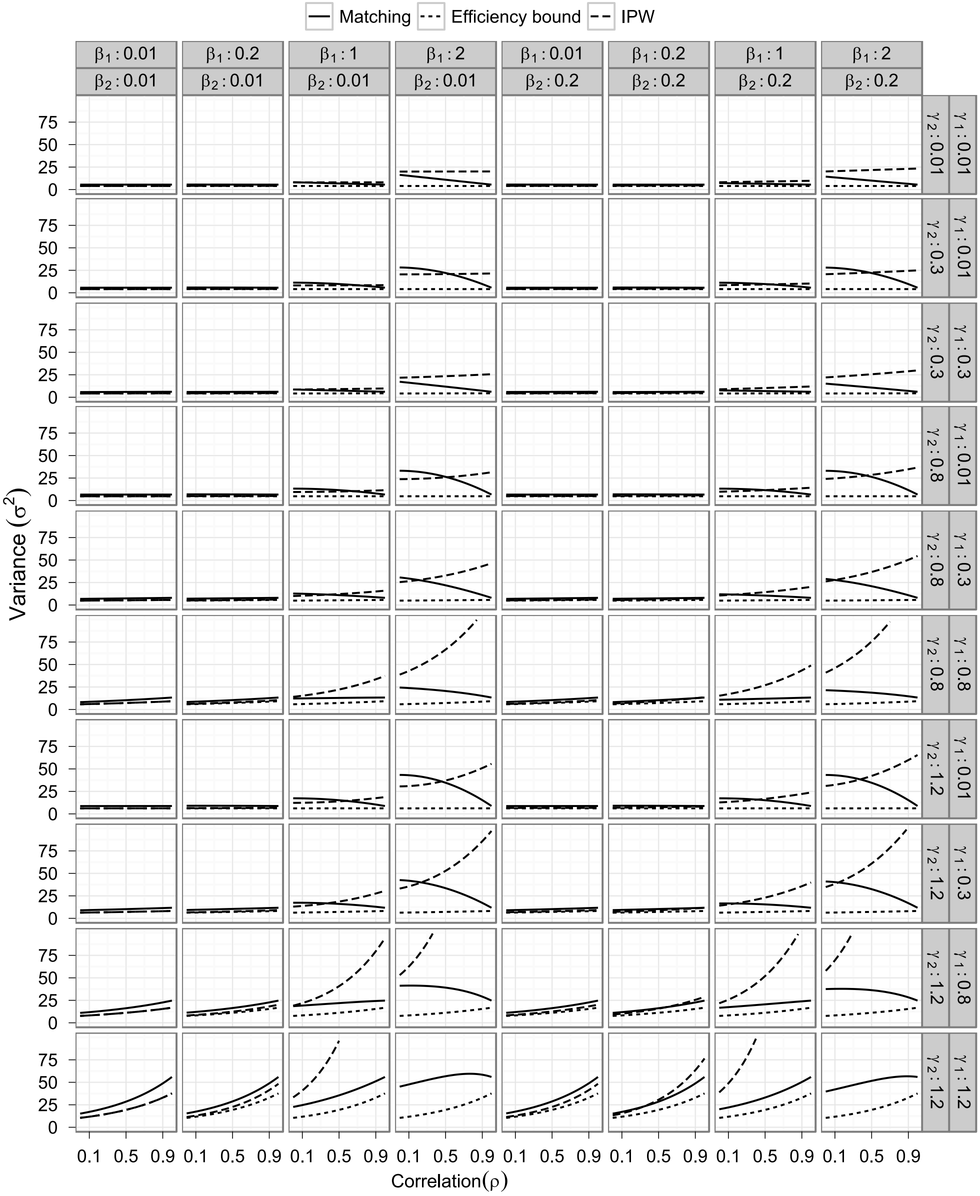
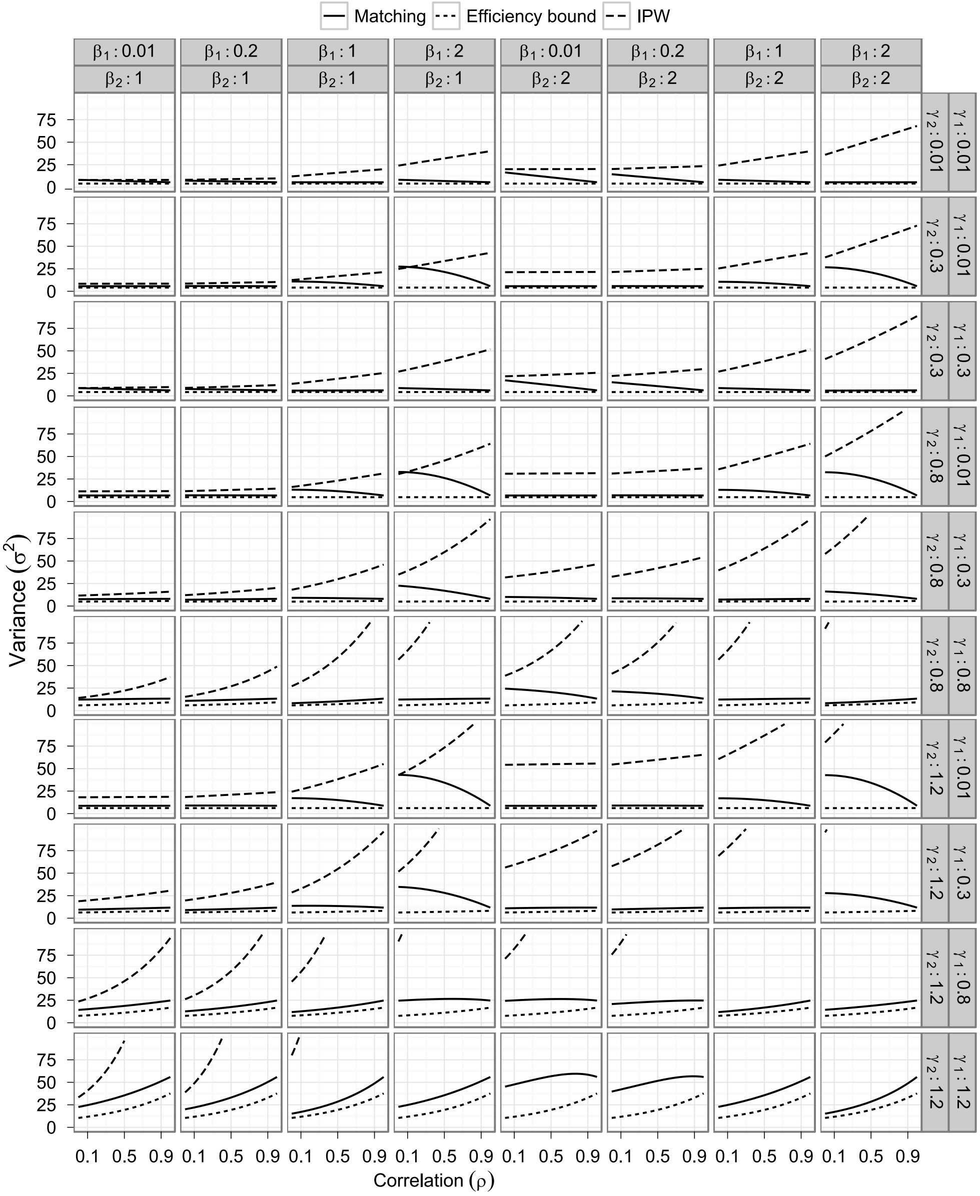


Figure 2: The relationship between ρ and σ^2_{IPW} and σ^2_{M}. Note that the subscript w is dropped for clarity.



Matching
Efficiency bound
IPW
β1 : 0.01
β1 : 0.2
β1 : 1
β1 : 2
β2 : 1
β2 : 2
γ1 : 0.01
γ1 : 0.3
γ1 : 0.8
γ1 : 1.2
γ2 : 0.01
γ2 : 0.3
γ2 : 0.8
γ2 : 1.2
Variance (σ2)
Correlation (ρ)

Interpreting Example 3 in terms of covariates, we see that a decrease in the variance of the matching estimator seems likely when one of the covariates is strongly related to the treatment assignment and weakly related to the outcome, while the other one is weakly related to the treatment assignment but strongly related to the outcome. This, however, does not imply that the variance of the matching estimator is smaller than the variance of the IPW estimator. In fact, it could be exactly the opposite. For instance, looking at the case when $\gamma_1 = 0.01$, $\gamma_2 = 1.2$, $\beta_{w1} = 2$ and $\beta_{w2} = 0.01$ it is apparent that when $\rho = 0$ the IPW estimator is more efficient than the matching estimator, and is actually close to the efficiency bound. Despite the negative slope of the variance, the matching estimator remains less efficient than the IPW estimator until the correlation between the covariates approaches 0.5.

Moreover, what also is apparent in Example 3 is that the IPW estimator in many cases is more sensitive to a change in the correlation, particularly when the overall relation between the covariates and the treatment assignment initially is strong.

3.3 The relative efficiency of the estimators

In the following we will examine the relative efficiency of the estimators by studying

$$R = \frac{\sigma^2_{\text{M}}}{\sigma^2_{\text{IPW}}}. \tag{18}$$

We begin with establishing a proposition regarding (18) that emphasizes some properties concerning the relationship between σ^2_{M} and σ^2_{IPW}.

Proposition 3 *Under Assumptions $1-5(ii)$ the maximum value of R in (18) is $1+1/2M$ which occurs when $\sigma_{Y_w,Z} = 0$. The minimum value of R is 0, which occurs at $\sigma^2_{Y_w,Z} - \sigma^2_Y \sigma^2_Z = 0$.*

To begin with, Proposition 3 shows that the maximum loss with respect to efficiency is smaller when selecting $\hat{\tau}_{\text{M}}$ instead of $\hat{\tau}_{\text{IPW}}$. Under Assumptions $1 - 5(ii)$, and when using one match only, the variance of the matching estimator is at most fifty percent greater than the variance of the IPW estimator. In the case of zero correlation between the covariates, this occurs when the parameter vector for the treatment assignment and the parameter vector for the outcome are orthogonal, i.e. $\beta_w \gamma = 0$. For R to be maximized when correlation is introduced orthogonality must still be maintained in the sense that $\beta_w \Sigma \gamma = 0$.

We now turn to how correlation affects the relative efficiency of the estimators.

Proposition 4 *Suppose that Assumptions $1-5(i-iii)$ hold. Then*

$$
\begin{aligned}
\frac{\partial R}{\partial \rho} =& \big\{2\sigma_{Y_w,Z}\,(1+4M)\,\big(\sigma_{Y_w,Z}\gamma' A\gamma\sigma^2_{Y_w} + \beta'_w A\beta_w\sigma_{Y_w,Z}\sigma^2_Z - 2\beta'_w A\gamma\sigma^2_{Y_w}\sigma^2_Z\big) \\
&+ 4\sigma_{Y_w,Z}e^{\sigma^2_Z}\,(1+2M)\,\big(\sigma^3_{Y_w,Z}\gamma' A\gamma - 2\beta'_w A\gamma\sigma^2_{Y_w}\sigma^2_Z(1+\sigma^2_Z) \\
&+ \sigma_{Y_w,Z}(\gamma' A\gamma\sigma^2_{Y_w} + \beta'_w A\beta_w\sigma^2_Z(1+\sigma^2_Z))\big) \\
&+ e^{\frac{1}{2}\sigma^2_Z}\,\big(\gamma' A\gamma\sigma^4_{Y_w}\sigma^4_Z + \sigma^4_{Y_w,Z}\gamma' A\gamma(1+4M)(2+\sigma^2_Z) \\
&- 4\beta'_w A\gamma\sigma_{Y_w,Z}\sigma^2_{Y_w}\sigma^2_Z(3+\sigma^2_Z+4M(2+\sigma^2_Z)) \\
&-\sigma^2_{Y_w,Z}(\gamma' A\gamma\sigma^2_{Y_w}(\sigma^2_Z-2) - 2\beta'_w A\beta\sigma^2_Z)(3+\sigma^2_Z+4M(2+\sigma^2_Z))\big)\big\} \\
&/\Big(8M(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}(\sigma^2_{Y_w,Z}+\sigma^2_{Y_w}))^2\sigma^4_Z\Big), \qquad (19)
\end{aligned}
$$

where A is a $k \times k$ hollow matrix with all diagonal elements equal to zero and all other elements equal to one.

Example 4 *For the setup as described in Example 2 we evaluate the derivative in (19). Figure 3 displays that $\partial R/\partial\rho$ is negative only for a small number of cases when σ_{Y_w} is very small.*

Under the studied conditions, the interpretation of Example 3 is that an increase in the correlation between covariates is for almost all cases more beneficial for the matching estimator relative to the IPW estimator. Hence, even if the variance of the IPW estimator is smaller than the variance of the matching estimator, increasing the correlation will decrease the difference between the variances.

4 Discussion

The objective of this study was to investigate the impact of correlation on the large sample variances of a matching estimator and an IPW estimator with all else being equal. This is motivated by the absence of theoretical results concerning how correlation among the covariates influence the variance of an estimator of the average causal effect. In this paper we show, under general model assumptions, how correlation is present in components of the asymptotic variances in a complex form. Thus, concluding that correlation has no effect on estimators of average treatment effects is not correct. In this regard, referring to standard consequences of multicollinearity is not sufficient. A main result of the paper is that the impact of the correlation differs between the two estimators. For the same conditions the variance of the respective estimators can change in

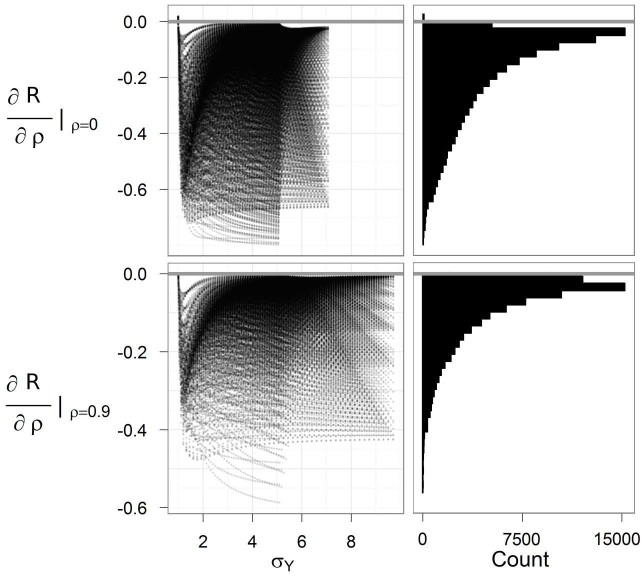


Figure 3: The relationship between σ_{Y_w} and $\partial R/\partial\rho$ evaluated at $\rho = 0$ and $\rho = 0.9$.

opposite directions, and the estimators display different degrees of sensitivity to a change in the correlation. Moreover, the strength of the confounding towards the outcome and the treatment, as illustrated in Figure 2, plays an important role which can in turn be seen as similar to the findings in Brookhart, Schneeweiss, Rothman, Glynn, Avorn, and Stürmer (2006). They conclude that including a variable that is strongly related to the treatment assignment but only weakly related to the outcome may increase the variance of a propensity score based estimator. Our results concerning the matching estimator are related, since if a variable is weakly related to the treatment assignment, introducing correlation with a variable that is strongly related to outcome will in turn increase the degree of dependence between the first variable and the outcome.

In practice a data analyst will face decisions concerning variable selection and/or transformations of the covariate vector. The current study motivates further investigations concerning the impact of correlation on causal effects estimators to assist empirical scientists to use strategies improving efficiency.

A Appendix

A.1 Proofs

Before proving the propositions, we introduce some useful results.

Begin with recalling Assumptions $2-4$ and their definitions. Then standard results tell us that Y_w and $Z=\gamma' X$ have a bivariate normal distribution with mean vector μ and covariance matrix Λ where

$$\mu = \begin{pmatrix} \mu_{Y_w} \\ \mu_Z \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} \sigma^2_{Y_w} & \sigma_{Y_w,Z} \\ \sigma_{Y_w,Z} & \sigma^2_Z \end{pmatrix} = \begin{pmatrix} \beta_w'\Sigma\beta_w + \sigma^2_{\varepsilon_w} & \beta_w'\Sigma\gamma \\ & \gamma'\Sigma\gamma \end{pmatrix}. \tag{A.1}$$

Further, consider a $k \times k$ equi-correlation matrix Σ with correlation ρ. A description of the properties of such a matrix can be found in Mardia, Kent, and Bibby (1979). For simplicity we merely note that $\rho \geq 0$ ensures a positive definite covariance matrix for all k. For a matrix A, $\text{vec}(A)$ is the vector generated by stacking all of the columns of A, the second below the first and so on. Thus, if $A \in \Re^{k\times p}$, then $\text{vec}(A) \in \Re^{kp\times 1}$. Now defining $A \in \Re^{k\times k}$ to be a hollow matrix with all off-diagonal elements equal to one, and making use of results found in e.g. Abadir and Magnus (2005) we get that

$$\begin{aligned} d\beta_w'\Sigma\beta_w =& \beta_w'(\text{d}\Sigma)\beta_w \\ =& (\beta_w' \otimes \beta_w')\,\text{dvec}\Sigma \\ =& (\beta_w' \otimes \beta_w')\,\text{vec}A \\ =& \text{vec}\,(\beta_w' A\beta_w) \\ =& \beta_w' A\beta_w \end{aligned} \tag{A.2}$$

where $\otimes$ denotes the Kronecker product.

Finally, we note that basic matrix calculus (Mardia, Kent, and Bibby 1979) gives that

$$\frac{\partial\beta_w'\Sigma\beta_w}{\partial\beta_w} = 2\Sigma\beta_w \tag{A.3}$$

and that

$$\frac{\partial(\beta_w'\Sigma\gamma)^2}{\partial\beta_w} = 2\beta_w'\Sigma\gamma'\Sigma\gamma. \tag{A.4}$$

Proof of Proposition 1. In order to formulate (5) in terms of the model defined in Assumptions $1-3$, we need some expectations. First note that if Z has a normal distribution, then e^Z and e^{-Z} are log-normal random variables with $E\left(e^Z\right) = e^{\mu_Z + \frac{1}{2}\sigma^2_Z}$ and $E\left(e^{-Z}\right) = e^{-\mu_Z + \frac{1}{2}\sigma^2_Z}$. Next, consider

$$\begin{aligned} E(Ze^Z) &= \int_{-\infty}^{\infty} \frac{ze^z}{\sqrt{2\pi\sigma^2_Z}} \exp\left[-\frac{(z-\mu_Z)^2}{2\sigma^2_Z}\right] dz \\ &= \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi\sigma^2_Z}} \exp\left[-\frac{(z^2 - 2(\mu_Z+\sigma^2_Z)z + (\mu_Z+\sigma^2_Z)^2 - (\mu_Z+\sigma^2_Z)^2 + \mu^2_Z)}{2\sigma^2_Z}\right] dz \\ &= \exp\left[\mu_Z + \frac{\sigma^2_Z}{2}\right] \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi\sigma^2_Z}} \exp\left[-\frac{(z-(\mu_Z+\sigma^2_Z))^2}{2\sigma^2_Z}\right] dz. \end{aligned}$$

Since this integral is the expected value of a normal distribution with mean $\mu_Z + \sigma^2_Z$ and variance σ^2_Z it follows that the expectations of this product and similar products are

$$E(Ze^{Z}) = \exp\left[\mu_Z + \frac{\sigma_Z^2}{2}\right]\left(\mu_Z + \sigma_Z^2\right)$$
$$E(Ze^{-Z}) = \exp\left[-\mu_Z + \frac{\sigma_Z^2}{2}\right]\left(\mu_Z - \sigma_Z^2\right)$$
$$E(Z^2e^{Z}) = \exp\left[\mu_Z + \frac{\sigma_Z^2}{2}\right]\left(\sigma_Z^2 + (\mu_Z + \sigma_Z^2)^2\right)$$
$$E(Z^2e^{-Z}) = \exp\left[-\mu_Z + \frac{\sigma_Z^2}{2}\right]\left(\sigma_Z^2 + (-\mu_Z + \sigma_Z^2)^2\right).$$

Using the previously defined expectations we find that

$$\begin{aligned}
E\left(Y_w e^{Z}\right) =& E\left[e^{Z}E(Y_w|Z)\right] \\
=& E\left[e^{Z}\left(\mu_{Y_w} + \frac{\sigma_{Y_w,Z}}{\sigma_Z^2}(Z-\mu_Z)\right)\right] \\
=& e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\mu_{Y_w} + \frac{\sigma_{Y_w,Z}}{\sigma_Z^2}E\left[e^{Z}Z\right] - \frac{\sigma_{Y_w,Z}}{\sigma_Z^2}e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\mu_Z \\
=& e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w} + \sigma_{Y_w,Z}\right) \\
E\left(Y_w e^{-Z}\right) =& e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w} - \sigma_{Y_w,Z}\right).
\end{aligned}$$

In a similar fashion we may calculate the expectations

$$\begin{aligned}
E\left(Y_w^2 e^{Z}\right) =& E\left[e^{Z}E(Y_w^2|Z)\right] \\
=& E\left[e^{Z}\left(\sigma_{Y_w}^2 - \frac{\sigma_{Y_w,Z}^2}{\sigma_Z^2} + \left[\mu_{Y_w} + \frac{\sigma_{Y_w,Z}}{\sigma_Z^2}(Z-\mu_Z)\right]^2\right)\right] \\
=& e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w}^2 + \sigma_{Y_w}^2 + \sigma_{Y_w,Z}^2 + 2\mu_{Y_w}\sigma_{Y_w,Z}\right) \\
E\left(Y_w^2 e^{-Z}\right) =& e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w}^2 + \sigma_{Y_w}^2 + \sigma_{Y_w,Z}^2 - 2\mu_{Y_w}\sigma_{Y_w,Z}\right).
\end{aligned}$$

Using Assumption 3, we rewrite (5)

$$\begin{aligned}
\sigma_{\text{IPW}}^2 =& E(Y_1^2 + e^{-Z}Y_1^2 - 2e^{-Z}\mu_{Y_1}Y_1 - 2\mu_{Y_1}Y_1 + \mu_{Y_1}^2 + e^{-Z}\mu_{Y_1}^2 \\
&+ Y_0^2 + e^{Z}Y_0^2 - 2e^{Z}\mu_{Y_0}Y_0 - 2\mu_{Y_0}Y_0 + e^{Z}\mu_{Y_0}^2 + \mu_{Y_0}^2).
\end{aligned}$$

Inserting $E\left(Y_w e^{Z}\right)$, $E\left(Y_w e^{-Z}\right)$, $E\left(Y_w^2 e^{Z}\right)$ and $E\left(Y_w^2 e^{-Z}\right)$ into σ_{IPW}^2 above and simplifying we finally get that

$$\sigma_{\text{IPW}}^2 = \sigma_{Y_1}^2 + e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_1}^2 + \sigma_{Y_1,Z}^2\right) + \sigma_{Y_0}^2 + e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_0}^2 + \sigma_{Y_0,Z}^2\right). \tag{A.5}$$

■

Proof of Proposition 2. For simplicity, denote the propensity score P. By Assumption 3, the inverse transformation of the propensity score is $Z = \log\left(P/(1-P)\right) = \text{logit}(P)$. Also, let $Y_w = Y_w$. The Jacobian and its determinant are

$$J = \begin{vmatrix} 1 & 0 \\ 0 & \frac{1}{p-p^2} \end{vmatrix} = \frac{1}{p-p^2}.$$

Thus the joint pdf of Y_w and P is

$$g(y_w,p) = \left|\frac{1}{p-p^2}\right| \frac{1}{2\pi\sigma_{Y_w}\sigma_Z\sqrt{1-\rho^2_{Y_wZ}}}$$
$$\times \exp\left[\frac{(y-\mu_{Y_w})^2}{2(1-\rho^2_{Y_wZ})\sigma^2_{Y_w}} - \frac{(\text{logit}(p)-\mu_Z)^2}{2(1-\rho^2_{Y_wZ})\sigma^2_Z} + \frac{2\rho_{Y_wZ}\,(y-\mu_Z)\,(\text{logit}(p)-\mu_Z)}{2(1-\rho^2_{Y_wZ})\sigma_{Y_w}\sigma_Z}\right].$$

Similarily, the pdf of P is using the change-of-variable technique

$$f(p) = \left|\frac{1}{p-p^2}\right| \frac{1}{\sqrt{2\pi\sigma^2_Z}} \exp\left[-\frac{(\text{logit}(p)-\mu_Z)^2}{2\sigma^2_Z}\right], \;\; 0 \le p \le 1.$$

Using these two distributions, we have the conditional distribution of potential outcome w given the propensity score, $g(y_w,p)/f(p)$. The factor $(p-p^2)^{-1}$ cancels out and what remains is the probability density function of a conditional normal distribution with mean

$$\mu_{Y_W} + \frac{\sigma_{Y_w,Z}}{\sigma^2_Z}\,[\text{logit}(p)-\mu_Z]$$

and variance

$$\sigma^2_{Y_w} - \frac{(\sigma_{Y_w,Z})^2}{\sigma^2_Z}.$$

By Assumption $1(i)$ we have the equality $E\,[Y_w|p(X)] = E\,[Y|W=w,p(X)]$. Hence, using the mean of the conditional distribution the first term in (8) becomes

$$\begin{aligned}
&E\,[(E(Y|W=1,p(X)) - E(Y|W=0,p(X)) - \tau)]^{\,2} \\
=&E\,[(E(Y_1|p(X)) - E(Y_0|p(X)) - \tau)]^{\,2} \\
=&E\left\{\left[\mu_{Y_1} + \frac{\sigma_{Y_1,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right]^2 + \left[\mu_{Y_0} + \frac{\sigma_{Y_0,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right]^2\right. \\
&\quad + \tau^2 + 2\tau\left[\mu_{Y_0} + \frac{\sigma_{Y_0,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right] - 2\tau\left[\mu_{Y_1} + \frac{\sigma_{Y_1,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right] \\
&\quad \left. -2\left[\mu_{Y_1} + \frac{\sigma_{Y_1,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right]\left[\mu_{Y_0} + \frac{\sigma_{Y_0,Z}}{\sigma^2_Z}\,(Z-\mu_Z)\right]\right\} \\
=&\frac{\sigma^2_{Y_1,Z}}{\sigma^2_Z} + \frac{\sigma^2_{Y_0,Z}}{\sigma^2_Z} - 2\frac{\sigma_{Y_1,Z}\sigma_{Y_0,Z}}{\sigma^2_Z}
\end{aligned}$$

The covariance matrix of the conditional distribution provides us with $V(Y|W=1,p(X))$ and $V(Y|W=0,p(X))$ in (8) and when again making use of the unconfoundedness assumption we get

$$V\,[Y|W=w,p(X)] = V\,[Y_w|p(X)] = \sigma^2_{Y_w} - \frac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}.$$

Using the results above, recalling that e^Z is a log-normal random variable, and substituting into (8) we have after simplification that

$$\sigma^2_{\mathrm{M}} = \frac{\sigma^2_{Y_1,Z} + \sigma^2_{Y_0,Z} - 2\sigma_{Y_1,Z}\sigma_{Y_0,Z}}{\sigma^2_Z} + \left(\sigma^2_{Y_1} - \frac{\sigma^2_{Y_1,Z}}{\sigma^2_Z}\right) E\left[1 + \frac{e^{-Z}(1+2M)}{2M} + \frac{1}{2M(1+e^Z)}\right] + \left(\sigma^2_{Y_0} - \frac{\sigma^2_{Y_0,Z}}{\sigma^2_Z}\right) E\left[\frac{2M+1}{2M} + \frac{e^{Z}(1+2M)}{2M} - \frac{1}{2M(1+e^Z)}\right]. \quad \text{(A.6)}$$

■

Proof of Corollary 1. Observe that $E\left(1+e^Z\right)^{-1}$ has no closed form expression, but through Assumption $5(ii)$ the distribution $h(Z) = \left(1+e^Z\right)^{-1}$ is symmetric around 0.5. Further, since $h(Z)$ is bounded for all real numbers, the expectation exists and we can conclude that $E(h(Z)) = 0.5$. Adding Assumption $5(i)$, leaving the choice of $w \in \{0,1\}$ arbitrary then allows us to write

$$\sigma^2_{\mathrm{M}} = \left(\sigma^2_{Y_w} - \frac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)\left(\frac{1+4M+(2+4M)e^{\frac{1}{2}\sigma^2_Z}}{2M}\right). \quad \text{(A.7)}$$

Adding Assumption $5(iii)$ enables us to use (A.2) by which we can show that

$$\frac{\partial\sigma^2_{\mathrm{M}}}{\partial\rho} = \frac{\left(1+4M+(2+4M)e^{\frac{1}{2}\sigma^2_Z}\right)\left(\beta'_w A\beta_w + \frac{\sigma^2_{Y_w,Z}\gamma' A\gamma}{(\sigma^2_Z)^2} - \frac{2\beta'_w A\gamma\sigma_{Y_w,Z}}{\sigma^2_Z}\right)}{2M} + \left(\sigma^2_{Y_w} - \frac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)\frac{(2+4M)\gamma' A\gamma e^{\frac{1}{2}\sigma^2_Z}}{4M} \quad \text{(A.8)}$$

where A is as defined in (A.2). ■

Remark: Note that if all elements in γ are equal then (A.8) becomes

$$\frac{\partial\sigma^2_{\mathrm{M}}}{\partial\rho} = -\frac{1}{k}\left(\sum_{j=1}^{k-1}\sum_{j'=j+1}^{k}(\beta_{wj}-\beta_{wj'})^2\right)\left(\frac{1+4M+(2+4M)e^{\frac{k}{2}(\gamma_j^2+\gamma_j^2(k-1)\rho)}}{2M}\right) - \left(\frac{2M+1}{2M}\right)(k-1)e^{\frac{k}{2}(\gamma_j^2+\gamma_j^2(k-1)\rho)}\gamma_j^2\left(-k+\sum_{j=1}^{k-1}\sum_{j'=j+1'}^{k}(\beta_{wj}-\beta_{wj'})^2(\rho-1)\right),$$

which can be reformulated as

$$\frac{\partial\sigma^2_{\mathrm{M}}}{\partial\rho} = f(M,k,\gamma_j,\rho) + g(M,k,\gamma_j,\rho)\sum_{j=1}^{k-1}\sum_{j'=j+1}^{k}(\beta_{wj}-\beta_{wj'})^2 \quad \text{(A.9)}$$

where

$$f(M,k,\gamma_j,\rho) = k\left(\frac{2M+1}{2M}\right)(k-1)e^{\frac{k}{2}(\gamma_j^2+\gamma_j^2(k-1)\rho)}\gamma_j^2$$

and

$$g(M,k,\gamma_j,\rho) = -\frac{1}{k}\left(\frac{1+4M+(2+4M)e^{\frac{k}{2}(\gamma_j^2+\gamma_j^2(k-1)\rho)}}{2M}\right) - \left(\frac{2M+1}{2M}\right)(k-1)e^{\frac{k}{2}(\gamma_j^2+\gamma_j^2(k-1)\rho)}\gamma_j^2(\rho-1).$$

Proof of Proposition 3. By Assumptions $5(i{-}ii)$, we have that $\sigma^2_{\text{IPW}} = 2\left(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}\left(\sigma^2_{Y_w} + \sigma^2_{Y_w,Z}\right)\right)$, where the choice of $w \in \{0,1\}$ is arbitrary. Together with (A.7), $R = \sigma^2_{\text{M}}/\sigma^2_{\text{IPW2}}$ can be written as

$$R = \frac{\left(\sigma^2_{Y_w} - \dfrac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)\left(\dfrac{1+4M+(2+4M)\,e^{\frac{1}{2}\sigma^2_Z}}{2M}\right)}{2\left(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}\left(\sigma^2_{Y_w} + \sigma^2_{Y_w,Z}\right)\right)}. \tag{A.10}$$

Recall that $\sigma^2_{Y_w} = \beta'_w \Sigma \beta_w + \sigma^2_{\varepsilon_w}$ and $\sigma_{Y_w,Z} = \beta'_w \Sigma \gamma$. Then using (A.3) and (A.4) we can easily show that

$$\frac{\partial \sigma^2_{\text{M}}}{\partial \beta_w} = \left(2\Sigma\beta_w - \frac{2\Sigma\gamma\sigma_{Y_w,Z}}{\sigma^2_Z}\right)\left(\frac{1+4M+(2+4M)\,e^{\frac{1}{2}\sigma^2_Z}}{2M}\right)$$

and that

$$\frac{\partial \sigma^2_{\text{IPW}}}{\partial \beta_w} = 4\left(\Sigma\beta_w + e^{\frac{1}{2}\sigma^2_Z}\left(\Sigma\beta_w + \Sigma\gamma\sigma_{Y_w,Z}\right)\right).$$

Using these derivatives and simplifying we get that

$$\begin{aligned}
\frac{\partial R}{\partial \beta_w} &= \frac{\dfrac{\partial \sigma^2_{\text{M}}}{\partial \beta_w}\sigma^2_{\text{IPW}} - \sigma^2_{\text{M}}\dfrac{\partial \sigma^2_{\text{IPW}}}{\partial \beta_w}}{\left(\sigma^2_{\text{IPW}}\right)^2} \\
&= \frac{\left(\dfrac{1+4M+(2+4M)\,e^{\frac{1}{2}\sigma^2_Z}}{2M}\right)\sigma_{Y_w,Z}\left(\sigma_{Y_w,Z}\Sigma\beta'_w - \sigma^2_{Y_w}\Sigma\gamma\right)\left(1 + e^{\frac{1}{2}\sigma^2_Z} + e^{\frac{1}{2}\sigma^2_Z}\sigma^2_Z\right)}{\sigma^2_Z\left(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}\left(\sigma^2_{Y_w} + \sigma^2_{Y_w,Z}\right)\right)^2}.
\end{aligned} \tag{A.11}$$

Thus local extrema occurs at $\sigma_{Y_w,Z} = 0$ and at $\sigma_{Y_w,Z}\Sigma\beta_w - \sigma^2_{Y_w}\Sigma\gamma = 0$. Keeping σ^2_Z fixed and observing the inequality

$$\begin{aligned}
R &= \frac{\left(\dfrac{1+4M+(2+4M)\,e^{\frac{1}{2}\sigma^2_Z}}{2M}\right)\left(\sigma^2_{Y_w} - \dfrac{\sigma^2_{Y_w,Z}}{\sigma^2_Z}\right)}{2\left(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}\left(\sigma^2_{Y_w} + \sigma^2_{Y_w,Z}\right)\right)} \\
&\le \frac{\left(\dfrac{1+4M+(2+4M)\,e^{\frac{1}{2}\sigma^2_Z}}{2M}\right)}{2\left(1 + e^{\frac{1}{2}\sigma^2_Z}\right)} = R|_{\sigma_{Y_w,Z}=0}
\end{aligned}$$

it is clear that $\sigma_{Y_w,Z} = 0$ maximizes R since a small change in β_w would add a negative term in the numerator and a positive term in the denominator, leading to a decrease in R. From this follows the important result

$$R|_{\sigma_{Y_w,Z}=0} \to 1 + \frac{1}{2M} \text{ as } \sigma^2_Z \to \infty,$$

which is the maximum value that R may attain. To evaluate the second extreme point, $\sigma_{Y_w,Z}\Sigma\beta_w - \sigma^2_{Y_w}\Sigma\gamma = 0$, we left multiply with γ' giving $\sigma^2_{Y_w,Z} - \sigma^2_{Y_w}\sigma^2_Z = 0$. Of course, $R|_{\sigma^2_{Y_w,Z} - \sigma^2_{Y_w}\sigma^2_Z = 0} = 0$. ■

Proof of Proposition 4. Again, under Assumptions $1-5(i-ii)$

$$\sigma^2_{\text{IPW}} = 2\left(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}\left(\sigma^2_{Y_w} + \sigma^2_{Y_w,Z}\right)\right).$$

Then, with the results in (A.2) we find that under Assumptions $1-5(i-iii)$

$$\begin{aligned}\frac{\partial\sigma^2_{\text{IPW}}}{\partial\rho} =& 2\Big(\beta_w' A\beta_w + (\beta_w' A\beta + 2\beta_w' A\gamma\sigma_{Y_w,Z})e^{\frac{1}{2}\sigma^2_Z} \\ &+ \frac{1}{2}e^{\frac{1}{2}\sigma^2_Z}\gamma' A\gamma(\sigma^2_{Y_w,Z} + \sigma^2_{Y_w})\Big).\end{aligned}$$

Thus, in order to find $\partial R/\partial\rho$ we merely use these two results, combine with previous results in (A.8) and apply the quotient rule. After some simplification we have that

$$\begin{aligned}\frac{\partial R}{\partial\rho} =& \big\{2\sigma_{Y_w,Z}\left(1+4M\right)\left(\sigma_{Y_w,Z}\gamma' A\gamma\sigma^2_{Y_w} + \beta_w' A\beta_w\sigma_{Y_w,Z}\sigma^2_Z - 2\beta_w' A\gamma\sigma^2_{Y_w}\sigma^2_Z\right) \\ &+ 4\sigma_{Y_w,Z}e^{\sigma^2_Z}\left(1+2M\right)\left(\sigma^3_{Y_w,Z}\gamma' A\gamma - 2\beta_w' A\gamma\sigma^2_{Y_w}\sigma^2_Z(1+\sigma^2_Z)\right. \\ &\left.+ \sigma_{Y_w,Z}(\gamma' A\gamma\sigma^2_{Y_w} + \beta_w' A\beta_w\sigma^2_Z(1+\sigma^2_Z)\right) \\ &+ e^{\frac{1}{2}\sigma^2_Z}\left(\gamma' A\gamma\sigma^4_{Y_w}\sigma^4_Z + \sigma^4_{Y_w,Z}\gamma' A\gamma(1+4M)(2+\sigma^2_Z)\right. \\ &- 4\beta_w' A\gamma\sigma_{Y_w,Z}\sigma^2_{Y_w}\sigma^2_Z(3+\sigma^2_Z+4M(2+\sigma^2_Z)) \\ &\left.-\sigma^2_{Y_w,Z}(\gamma' A\gamma\sigma^2_{Y_w}(\sigma^2_Z-2) - 2\beta_w' A\beta\sigma^2_Z)(3+\sigma^2_Z+4M(2+\sigma^2_Z))\right)\big\} \\ &/\left(8M(\sigma^2_{Y_w} + e^{\frac{1}{2}\sigma^2_Z}(\sigma^2_{Y_w,Z}+\sigma^2_{Y_w}))^2\sigma^4_Z\right). \qquad \text{(A.13)}\end{aligned}$$

■

A.2 Excluding combinations in Example 2 and Example 3

We illustrate the distribution of the propensity score for the treated and the controls for different values of ρ and γ. Here, we see that for small and large values of σ^2_Z we either have insufficient overlap or $f(p(X) \mid W=1) \approx f(p(X) \mid W=0)$, thus motivating the exclusion of combinations in Example 2 and Example 3 (see Figure A.2).

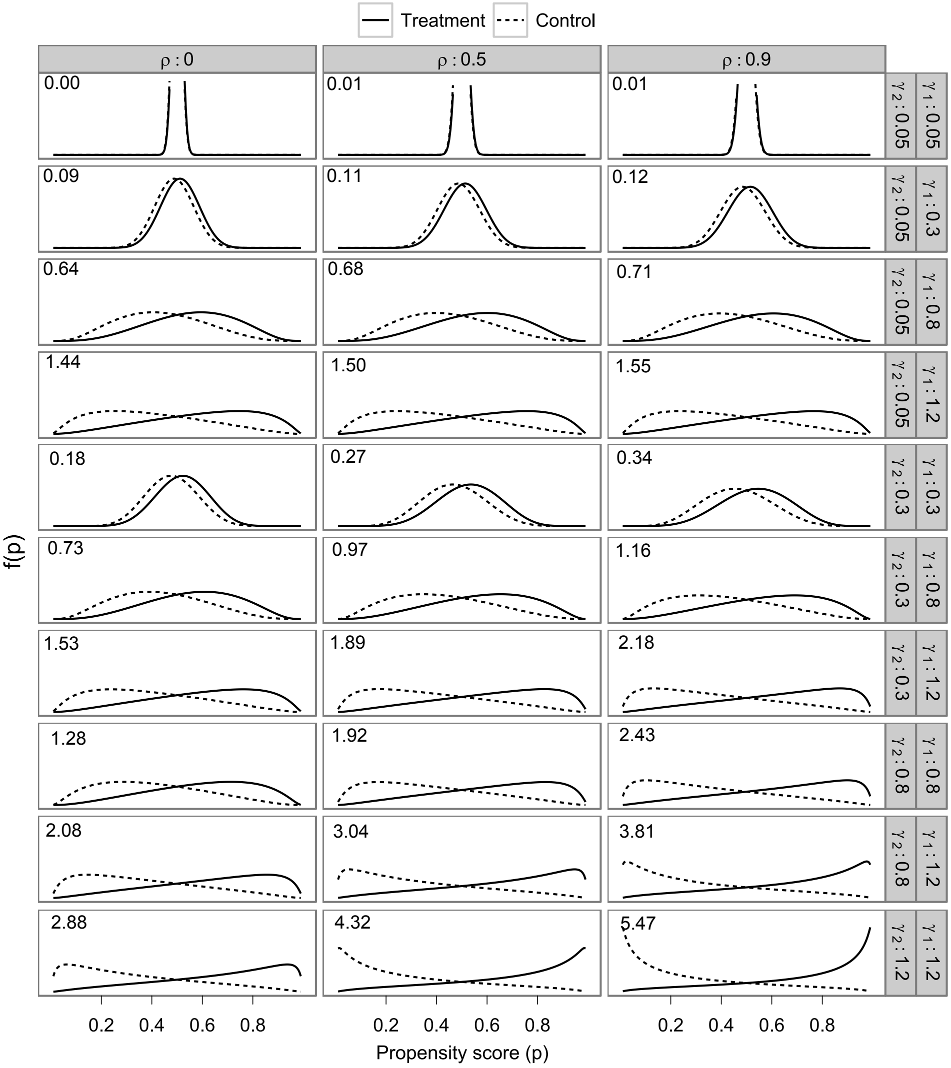


Figure 4: The distribution of the propensity score for different values of γ and ρ. The numbers in the upper left corners are σ_Z^2.

References

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

Abadir, K. and J. Magnus (2005). *Matrix Algebra*. Cambridge: Cambridge University Press.

Brookhart, M. A., S. Schneeweiss, K. K. Rothman, R. J. Glynn, J. Avorn, and T. Stürmer (2006). Variable selection for propensity score models. *American Journal of Epidemiology 163*, 1149–1156.

Dehejia, R. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association 94*, 1053–1062.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*, 315–331.

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

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*, 1161–1189.

Horvitz, D. and D. Thompson (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association 47*, 663–685.

Imbens, G. W., W. Newey, and G. Ridder (2005). Mean-squared error calculations for average treatment effects. *IEPR Working Paper*, 05:34, Institute of Economic Policy Research, University of Southern California.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*, 5–86.

Kang, J. D. Y. and J. L. Schafer (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science 22*, 523–539.

Lunceford, J. K. and M. Davidian (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine 23*, 2937–2960.

Luo, Z., J. C. Gardiner, and C. J. Bradly (2010). Applying propensity score methods in medical research: pitfalls and prospects. *Medical Care Research and Review 67*, 528–554.

Mardia, K., J. Kent, and J. Bibby (1979). *Multivariate Analysis*. London: Academic Press.

Millimet, D. L. and R. Tchernis (2009). On the specification of propensity scores, with applications to the analysis of trade policies. *Journal of Business and Economic Statistics 27*, 397–415.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych X*, 1–51. In Polish, English translation by D. Dabrowska and T. Speed in *Statistical Science*, **5**, 465–472, 1990.

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

Stuart, E. A. (2010). Matching methods for causal inference: A review and look forward. *Statistical Science 25*, 1–21.

Thoemmes, F. J. and E. S. Kim (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Statistics 46*, 90–118.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Waernbaum, I. (2010). Propensity score model specification for estimation of average treatment effects. *Journal of Statistical Planning and Inference 140*, 1948–1956.

Waernbaum, I. (2012). Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. *Statistics in Medicine 31*, 1572–1581.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: The MIT Press.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics 141*, 1281–1301.

Publication series published by IFAU – latest issues

Rapporter/Reports

2012:20 Ulander-Wänman Carin ”Flexibilitet – en dominerande diskurs i de anställningsvillkor som rör sysselsättningstrygghet för arbetstagare i kommun och landsting”

2012:21 Wikström Christina and Magnus Wikström ”Urval till högre utbildning – Påverkas betygens prediktionsvärde av ålder?

2012:22 Karimi Arizo, Erica Lindahl and Peter Skogman Thoursie ”Effekter av föräldrapenning på arbetsutbud”

2012:23 Sibbmark Kristina ”Arbetsmarknadspolitisk översikt 2011”

2012:24 Liljeberg Linus, Sara Martinson and Jonas Thelander ”Vad innebär det att bli coachad? En utvärdering av jobbcoachningen vid Arbetsförmedlingen”

2012:25 van den Berg Gerard J., Petter Lundborg and Johan Vikström ”De långsiktiga ekonomiska och sociala konsekvenserna av att förlora ett barn”

2012:26 Hägglund Pathric, Per Johansson and Lisa Laun ”Rehabiliteringsgarantin”

2012:27 Engdahl Mattias and Olof Åslund ”Ekonomiska drivkrafter och studieresultat – effekter av Sfi-bonus”

2012:28 Lindgren Karl-Oskar ”Arbetsplatsstorlek och sjukfrånvaro”

2013:1 Olsson Martin ”Anställningsskydd och föräldrarelaterad frånvaro”

2013:2 Angelov Nikolay, Per Johansson and Erica Lindahl ”Det envisa könsgapet i inkomster och löner – Hur mycket kan förklaras av skillnader i familjeansvar?”

2013:3 Vikman Ulrika ”Så påverkar föräldraförsäkringen nyanlända invandrares etablering på arbetsmarknaden”

Working papers

2012:20 Angelov Nikolay and Arizo Karimi ”Mothers’ income recovery after childbearing”

2012:21 Wikström Christina and Magnus Wikström ”University entrance selection and age at admission”

2012:22 Karimi Arizo, Erica Lindahl and Peter Skogman Thoursie ”Labour supply responses to paid parental leave”

2012:23 van den Berg Gerard J., Petter Lundborg and Johan Vikström ”The economics of grief”

2012:24 Åslund Olof and Mattias Engdahl ”The value of earning for learning: performance bonuses in immigrant language training”

2012:25 Collet Francois and Peter Hedström ”Endogenous tie formation mechanisms in a directed network generated by employee mobility”

2012:26 Lindgren Karl-Oskar ”Workplace size and sickness absence transitions”

2013:1 Nekby Lena, Peter Skogman Thoursie and Lars Vahtrik ”Examination behavior – Gender differences in preferences?”

2013:2 Olsson Martin “Employment protection and parental child care”

2013:3 Angelov Nikolay, Per Johansson and Erica Lindahl “Is the persistent gender gap in income and wages due to unequal family responsibilities?”

2013:4 Vikman Ulrika “Paid parental leave to immigrants: An obstacle to labor market entrance?”

2013:5 Pingel Ronnie and Ingeborg Waernbaum “Effects of correlated covariates on the efficiency of matching and inverse probability weighting estimators for causal inference”

Dissertation series

2012:1 Laun Lisa “Studies on social insurance, income taxation and labor supply”