Pingel, Ronnie; Waernbaum, Ingeborg

**Working Paper**

# Correlation and efficiency of propensity score-based estimators for average causal effects

Working Paper, No. 2015:3

**Provided in Cooperation with:**
IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

*Suggested Citation:* Pingel, Ronnie; Waernbaum, Ingeborg (2015) : Correlation and efficiency of propensity score-based estimators for average causal effects, Working Paper, No. 2015:3, Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala

This Version is available at:
https://hdl.handle.net/10419/129370

IFAU
Institute for Evaluation of Labour
Market and Education Policy

# Correlation and efficiency of propensity score-based estimators for average causal effects

Ronnie Pingel
Ingeborg Waernbaum

WORKING PAPER 2015:3

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

# Correlation and efficiency of propensity score-based estimators for average causal effects[a]

by

Ronnie Pingel[b] and Ingeborg Waernbaum[c]

12th February, 2015

**Abstract**

Propensity score based-estimators are commonly used to estimate causal effects in evaluation research. To reduce bias in observational studies researchers might be tempted to include many, perhaps correlated, covariates when estimating the propensity score model. Taking into account that the propensity score is estimated, this study investigates how the efficiency of matching, inverse probability weighting and doubly robust estimators change under the case of correlated covariates. Propositions regarding the large sample variances under certain assumptions of the data generating process are given. The propositions are supplemented by several numerical large sample and finite sample results from a wide range of models. The results show that the correlation may increase or decrease the variances of the estimators. There are several factors that influence how correlation affects the variance of the estimators, including the choice of estimator, the strength of the confounding towards outcome and treatment, and whether a constant or non-constant causal effect is present.

Keywords: doubly robust; inverse probability weighting; matching; observational study.
JEL-codes: C13; C40; C52

[a]We like to thank Per Johansson and Martin Huber for valuable comments. We also like to thank seminar participants at the joint IFAU and Department of Statistics seminar. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project p2010037. Ronnie Pingel acknowledge financial support from the Institute for Evaluation of Labour Market and Education Policy, Sweden.

[b]Department of Statistics, Uppsala University, ronnie.pingel@statistics.uu.se

[c]Department of Statistics, Umeå University and Institute for Evaluation of Labour Market and Education Policy, ingeborg.waernbaum@stat.umu.se

## Table of contents

# 1 Introduction

A natural starting point in the evaluation of a treatment is to compare average outcomes for treated and control units. When we have observational data, i.e. the assignment of treatment is not randomised, we need to adjust for differences in background variables, covariates, between the treated and controls. In causal inference this is sometimes referred to as estimation of treatment effects under unconfoundedness, no unmeasured confounding or selection on observables (see Imbens and Wooldridge, 2009, for a review). In general, a test for unconfoundedness cannot be made with the available data and subject matter theory provides the key guidance. Thus, for an empirical researcher it might seem reasonable to include many variables. The role of the covariates has been extensively discussed in the literature from different aspects. Covariate selection and optimal sets of covariates have been discussed both in parametric (Vansteelandt et al., 2010) and non-parametric settings (Hahn, 2004; de Luna et al., 2011; White and Lu, 2011). For propensity score models, simulation studies have been performed to investigate the effect of covariate selection (Brookhart et al., 2006) and "over-modeling" of the propensity score model by including higher-order terms (Millimet and Tchernis, 2009). Theoretical results on the inclusion of additional covariates have also been derived (Lunceford and Davidian, 2004). The studies by Brookhart et al. and Lunceford and Davidian share the same conclusion: adding extra information to the model brings about an increase in efficiency if the additional variables are related to the outcome, but not otherwise.

In this paper we study the effect of correlation between covariates in propensity score based-estimators. Correlation between variables that cause multicollinearity is a long-familiar problem in regression analysis. Multicollinearity renders unstable matrix inversion because the sizes of the numbers in the inverted matrix fluctuate wildly with only small changes in the sizes of the elements of the correlation matrix of the covariates. Propensity score-based estimators constitute a class of estimators that are widely used among empirical researchers (Stürmer et al., 2006; Connors et al., 1996) studying causal effects. It has been argued that multicollinearity does not affect the variance of an estimator of the average causal effect Stuart (2010). This is because the main concern is not with

the individual parameter estimates but with the predicted probabilities of the treatment assignment. Pingel and Waernbaum (2014) show how correlation among the covariates influences the large sample variance of a matching estimator and an inverse probability weighting (IPW) estimator using the true propensity score. In the present study we extend the results to estimated propensity scores . Here we investigate large and finite sample variances of three estimators of an average causal effect: matching, IPW and a doubly robust (DR) estimator. By assuming parametric models with normally distributed covariates, linear outcome models and a logistic regression for the propensity score, we give theoretical results on the effect of correlation on the efficiency. To generalise, we investigate a wider range of covariate distributions and outcome models showing numerical results that in most cases are in the same direction. Further, finite sample variances for all models are investigated in simulations.

Our study demonstrates that the efficiency of the DR estimator is only influenced by the correlation matrix through the multiplication of the parameter vector in the propensity score model. IPW and matching stand in contrast since the variances contain terms including also a quadratic form of the outcome and the scalar formed by multiplying the propensity score vector, the covariance matrix from the propensity score model, and the parameter vector for the outcome. This means that the magnitude of the covariates' influence on treatment assignment as well as on outcome interacts with the correlation in the components of the variance of the estimators. We show analytically and in simulation that the efficiency of the estimators are affected differently by the correlation.

In the next section we introduce the theoretical framework and notation, as well as define the different estimators (matching, IPW and DR). In Section 3 we describe the effect of correlated covariates in a regression setting when the causal effect is a regression coefficient. In Section 4 the propensity score-based estimators and their properties are described. The theoretical properties under restricted assumptions are presented in Section 5. In Section 6 the assumptions are relaxed and we give numerical results and perform simulation studies for a wider range of models. Section 7 contains an application and Section 8 concludes with a brief discussion.

## 2 Framework

Consider a random sample of $N$ units assigned to either a treatment group, $W = 1$, or a control group, $W = 0$, and that we wish to study how the treatment affects a response variable of interest. Following the Neyman-Rubin framework with potential outcomes (Neyman, 1923; Rubin, 1974), the unit-specific causal effect may be defined as $Y_1 - Y_0$. We need to also define the potential outcome means, $E(Y_1) = \mu_{Y_1}$, $E(Y_0) = \mu_{Y_0}$, and variances, $V(Y_1) = \sigma^2_{Y_1}$, $V(Y_0) = \sigma^2_{Y_0}$. The unit-specific effect is not estimable since only one of the two potential outcomes is realised for each unit. However, the aim of this study is not estimation of unit-specific effects, but estimation of the average causal (treatment) effect, $\tau = E(Y_1 - Y_0)$. The average causal effect is estimable under certain assumptions, which, for instance, follows from a randomised experiment. In the event that the treatment is not randomised and treatment assignment is affected by observed variables, $X$, the average causal effect can be identified under the assumption of strong ignorability (Rosenbaum and Rubin, 1983).

**Assumption 1 (Strong ignorability)** *(i) (Unconfoundedness)*
$(Y_1, Y_0) \perp\!\!\!\perp W \mid X$, *and (ii) (Overlap)* $0 < \Pr(W = 1|X) < 1$.

Furthermore, the stable unit treatment value assumption holds (Rubin, 1980).

Since Rosenbaum and Rubin's (1983) seminal paper, we now recognise that instead of conditioning on the covariates directly, it is sufficient to condition on the propensity score. In this paper the propensity score is formulated $\Pr(W = 1|X, \gamma) = p(Z) = (1 + e^{-Z})^{-1}$, where the logit $Z = X'\gamma$ includes $\gamma$, a $k$-dimensional parameter vector. Further, let $E(Z) = \mu_Z$ and $V(Z) = \sigma^2_Z$ and define the covariances $\mathrm{Cov}(Y_1, Z) = \sigma_{Y_1,Z}$ and $\mathrm{Cov}(Y_0, Z) = \sigma_{Y_0,Z}$. To impose an intercept we simply put the unit vector as the first covariate. Note that $p'(Z) = p(Z)[1 - p(Z)]$. When estimating $\gamma$ from a sample using maximum likelihood estimation the Fisher information matrix for $\gamma$ can then be formulated

$$I = E\left[p'(Z)XX'\right]. \tag{1}$$

See Lee (1990) for details. Under certain regularity conditions, $\hat{\gamma}$ is asymptotically normally distributed with the covariance matrix equal to the inverse of the information matrix.

**Assumption 2 (Propensity score model)** *The propensity score is generated and consistently estimated by a logistic regression model.*

Furthermore, we denote the covariance matrix of the covariates $\Sigma$. The corresponding correlation matrix is defined $(\rho_{st}) = [\text{diag}(\Sigma)]^{-1/2}\Sigma[\text{diag}(\Sigma)]^{-1/2}$, where $\text{diag}(\Sigma)$ is the diagonal matrix acquired by keeping the diagonal elements of $\Sigma$ and replacing all other elements with zero. The correlation between two covariates $s$ and $t$ is referred to as $\rho_{st}$. Note that for brevity, the index is sometimes omitted.

# 3 A regression estimator

Having established the framework in Section 2, we are able to clarify the issue of correlated covariates and how this adds to the discussion of what is usually referred to as multicollinearity. We do this by studying the ordinary least squares (OLS) estimator. Assume a standard bivariate normal distribution with variables $X_1$ and $X_2$ with correlation $\rho_{12}$. An outcome is generated by

$$Y = \tau W + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \;\; \varepsilon \sim N(0, \sigma_\varepsilon^2),$$

with the parameters $\beta_1$ and $\beta_2$, and the error term $\varepsilon$. Typically, in textbooks multicollinearity refers to the inflation of the variances $V(\hat{\beta}_1)$ and $V(\hat{\beta}_2)$ that is due to a large $\rho_{12}$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are least squares estimators of $\beta_1$ and $\beta_2$. However, interest lies in estimating the average causal effect $\tau$. Thus, the aim is to study how the variance $V(\hat{\tau})$ is affected by $\rho_{12}$, where in this case $\hat{\tau}$ is a least squares estimator of $\tau$. Consider the latent treatment variable

$$W^* = \gamma_1 X_1 + \gamma_2 X_2 + \eta, \;\; \eta \sim N(0, \sigma_\eta^2),$$

and the variance decomposition

$$V(W^*) = (\gamma_1^2 + \gamma_2^2 + 2\rho_{12}\gamma_1\gamma_2) + \sigma_\eta^2 = \text{explained variance} + \text{error variance}.$$

For fixed $\sigma_\eta^2$, an increase in $\rho_{12}$ will decrease or increase the explained variance, $R^{*2}$, depending on the signs of $\rho_{12}$, $\gamma_1$ and $\gamma_2$. Defining $Z^* = \gamma_1 X_1 + \gamma_2 X_2$, the explained variance can be written $R^{*2} = \text{cor}(W^*, Z^*)^2$.

However, in practice we only observe the binary treatment variable

$$W = 1[W^* > 0],$$

where $1[\cdot]$ denotes the indicator function. The variance of $\hat{\tau}$ is

$$V(\hat{\tau}) = \frac{\sigma_\varepsilon^2}{(1-R^2)\sum_{i=1}^N (W_i - \bar{W})^2},$$

where $\bar{W} = N^{-1}\sum_{i=1}^N W_i$ and $R^2$ is the explained variance of the observed variable. Observe that in this setting $E(W) = 0.5$, rendering $\sum_{i=1}^N (W_i - \bar{W})^2$ to be unaffected by $\rho_{12}$. This implies that an increase in $\rho_{12}$ will lead to an increase in $V(\hat{\tau})$ through $R^2$. However, because $W$ is binary we use the result in Cohen (1983) stating that for the case with two normally distributed variables, such as $W^*$ and $Z^*$, in which one is dichotomised into equally sized groups, $\text{cor}(W, Z^*) = 0.798 \cdot \text{cor}(W^*, Z^*)$. Thus, a change in $\rho_{12}$ affects the explained variance of the observed variable, $R^2$, in the same direction as $R^{*2}$, but to a lesser extent. Since the maximum value of $R^2$ is approximately 0.64, the variance will not be materially affected (i.e. inflated).

In this simple example we have established that an increase in the correlation between two variables influences the variance of an estimator of the average causal effect through the latent treatment variable structure. To our knowledge, this study and the study by Pingel and Waernbaum (2014) are the only to address and formalise this issue. In the remainder of this paper focus is not on the least squares estimator of the average causal effect but on how correlation affects the variance of estimators using the propensity score.

# 4 Propensity score-based estimators

We study three commonly used propensity score-based estimators: an IPW estimator, a DR estimator and a propensity score matching estimator. Under Assumptions 1–2, the estimators are consistent and approximately normally distributed in large samples. The studied IPW estimator is the normalised IPW estimator proposed by Hirano et al. (2003):

$$\hat{\tau}_{\text{IPW},\hat{p}} = \left( \sum_{i=1}^{N} \frac{W_i}{\hat{p}(Z_i)} \right)^{-1} \sum_{i=1}^{N} \frac{W_i Y_i}{\hat{p}(Z_i)} - \left( \sum_{i=1}^{N} \frac{1-W_i}{1-\hat{p}(Z_i)} \right)^{-1} \sum_{i=1}^{N} \frac{(1-W_i)Y_i}{1-\hat{p}(Z_i)}. \tag{2}$$

As described by Lunceford and Davidian (2004), the asymptotic variance of $\sqrt{N}(\hat{\tau}_{\text{IPW},\hat{p}} - \tau)$ is

$$\sigma^2_{\text{IPW},\hat{p}} = \sigma^2_{\text{IPW}} - a' I^{-1} a. \tag{3}$$

The first part of the variance expression is the asymptotic variance when using the true propensity score

$$\sigma^2_{\text{IPW}} = E\left[ \frac{(Y_1 - \mu_{Y_1})^2}{p(Z)} + \frac{(Y_0 - \mu_{Y_0})^2}{1-p(Z)} \right]. \tag{4}$$

The second part adjusts for the estimation of the propensity score and includes the $k$-dimensional vector

$$a = E\left[ \left( \frac{Y_1 - \mu_{Y_1}}{p(Z)} + \frac{Y_0 - \mu_{Y_0}}{1-p(Z)} \right) p'(Z) X \right]. \tag{5}$$

As an alternative to the IPW estimator we study the DR estimator (Robins et al. (1994); Lunceford and Davidian (2004)). Let $m_w(X, \beta_w) = E(Y|W = w, X)$ denote the regression of $Y$ on $X$ in group $w$ and let $\hat{\beta}_w$ be an estimator for the regression parameter $\beta_w$ using subjects within group $w$ only. The DR estimator is defined as

$$\begin{aligned} \hat{\tau}_{\text{DR}} = & \frac{1}{N} \sum_{i=1}^{N} \frac{W_i Y_i - (W_i - \hat{p}(Z_i))\, \hat{m}_1\left(X_i, \hat{\beta}_1\right)}{\hat{p}(Z_i)} \\ & - \frac{1}{N} \sum_{i=1}^{N} \frac{(1-W_i)Y_i + (W_i - \hat{p}(Z_i))\, \hat{m}_0\left(X_i, \hat{\beta}_0\right)}{1-\hat{p}(Z_i)}. \end{aligned} \tag{6}$$

If the propensity score model and the regression models are correctly specified, the large sample variance of $\sqrt{N}(\hat{\tau}_{\text{DR}} - \tau)$ is

$$\sigma^2_{\text{DR}} = \sigma^2_{\text{IPW}} - b, \tag{7}$$

where

$$b = E\left[\left(\sqrt{\frac{1-p(Z)}{p(Z)}}\left[E\left(Y_1|X\right) - \mu_{Y_1}\right] + \sqrt{\frac{p(Z)}{1-p(Z)}}\left[E\left(Y_0|X\right) - \mu_{Y_0}\right]\right)^2\right] \tag{8}$$

is a positive scalar. The large sample variance of the DR estimator is the same irrespective of using the known or the estimated propensity score (Lunceford and Davidian, 2004).

The third estimator we consider is a propensity score matching estimator with replacement (e.g., Abadie and Imbens, 2006). We define the distance between two units $i$ and $i'$ from opposite treatment groups $d_{ii'} = \mid \hat{p}(Z_i) - \hat{p}(Z_{i'}) \mid$. Thus, for each $i$ there is a set $\mathscr{I} = \{1, 2, \ldots, i', \ldots, M\}$ of indices of the $M$ individuals with the smallest order statistics $d_{i(i')}$, $i' \leq M$. The matching estimator matching treated and controls to a fixed number of $M$ matches can then be formulated

$$\hat{\tau}_{\text{M}} = \frac{1}{N}\sum_{i=1}^{N} W_i(Y_i - \hat{Y}_{0i}) + (1 - W_i)(\hat{Y}_{1i} - Y_i), \tag{9}$$

where $\hat{Y}_{0i} = \sum_{i' \in \mathscr{I}} Y_{i'}/M$ and $\hat{Y}_{1i} = \sum_{i' \in \mathscr{I}} Y_{i'}/M$ are the means of the observed response for the $M$ matched individuals. Abadie and Imbens (2012) show that the large sample variance of $\sqrt{N}(\hat{\tau}_{\text{M},\hat{p}} - \tau)$ using the estimated propensity score is

$$\sigma^2_{\text{M},\hat{p}} = \sigma^2_{\text{M}} - c' I^{-1} c. \tag{10}$$

The first part,

$$\begin{aligned}\sigma^2_{\text{M}} = &E\left[\left(E\left[Y|W=1,p(Z)\right]-E\left[Y|W=0,p(Z)\right]-\tau\right)^2\right]\\ &+E\left[V\left[Y|W=1,p(Z)\right]\left(\frac{1}{p(Z)}+\frac{1}{2M}\left(\frac{1}{p(Z)}-p(Z)\right)\right)\right]\\ &+E\left[V\left[Y|W=0,p(Z)\right]\left(\frac{1}{1-p(Z)}+\frac{1}{2M}\left(\frac{1}{1-p(Z)}-(1-p(Z))\right)\right)\right] \qquad (11)\end{aligned}$$

is the asymptotic variance of the matching estimator when using the true propensity score. The second part involves the $k$-dimensional vector

$$c = E\left[\left(\frac{\text{Cov}\left(X,Y|p(Z),W=1\right)}{p(Z)}+\frac{\text{Cov}\left(X,Y|p(Z),W=0\right)}{1-p(Z)}\right)p'(Z)\right] \qquad (12)$$

consisting of the weighted covariances between the covariates and the outcome conditional on the propensity score and treatment.

The advantages and disadvantages of the estimators have been described elsewhere (e.g., Lunceford and Davidian, 2004; Waernbaum, 2012). The main point is that although all three estimators utilise the propensity score, they do so differently. Both the IPW and DR estimator use the propensity score to reweight the data, creating a pseudo-population with missing potential outcomes. The matching estimator uses the propensity score as a balancing score, i.e. it imputes the unobserved potential outcome with the outcome on units sharing similar characteristics in the opposite treatment group. All estimators are easily implemented in practice, but the IPW estimator is sensitive when the propensity score is too close to zero or one. The DR estimator performs much better when both the outcome model part and the propensity score are correctly specified, but studies have shown that it is not efficient when the outcome model is wrong (Waernbaum, 2012).

# 5 Analytic results on the asymptotic variance

In this section we present analytical results for how the asymptotic variances of the estimators are affected by the correlation $\rho_{st}$. To emphasise the role of the intercept we choose not to include 1 in $X$ in the following assumption.

**Assumption 3** *(i) The covariates X follow a k-variate normal distribution with a zero mean vector and covariance matrix* $\Sigma$*. (ii) The potential outcomes are generated by* $Y_1 = \alpha_1 + X'\beta_1 + \varepsilon_1$ *and* $Y_0 = \alpha_0 + X'\beta_0 + \varepsilon_0$*, where* $\beta_1$*,* $\beta_0$ *are parameter vectors and* $\varepsilon_1, \varepsilon_0 \sim N(0, \sigma_\varepsilon^2)$*. The error terms are uncorrelated with each other and X. (iii) Constant causal effect (i.e.* $\beta_1 = \beta_0$*).*

Because of the linearity of the parameters, the variances and covariances of the potential outcomes and the logit are

$$\sigma_{Y_1}^2 = \beta_1'\Sigma\beta_1 + \sigma_\varepsilon^2, \;\; \sigma_{Y_0}^2 = \beta_0'\Sigma\beta_0 + \sigma_\varepsilon^2, \;\; \sigma_Z^2 = \gamma'\Sigma\gamma,$$

and

$$\sigma_{Y_1,Z} = \beta_1'\Sigma\gamma, \;\; \sigma_{Y_0,Z} = \beta_0'\Sigma\gamma.$$

**Proposition 1** *Under Assumptions 1–3, the asymptotic variances of* $\hat{\tau}_{IPW,\hat{p}}$*,* $\hat{\tau}_{DR}$*, and* $\hat{\tau}_{M,\hat{p}}$ *are*

$$\begin{aligned}
\sigma_{IPW,\hat{p}}^2 &= 2\sigma_{Y_w}^2 + e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_w}^2 + \sigma_{Y_w,Z}^2\right) + e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_w}^2 + \sigma_{Y_w,Z}^2\right) \\
&- \left[\frac{\left(\sigma_{Y_w}^2 - \sigma_\varepsilon^2\right)}{E\left[p'(Z)\right]} - \frac{\sigma_{Y_w,Z}^2}{E\left[p'(Z)\right]\sigma_Z^2} + \frac{\sigma_{Y_w,Z}^2 E\left[p'(Z)\right]}{E\left[p'(Z)\right]E\left[p'(Z)Z^2\right] - \left(E\left[p'(Z)Z\right]\right)^2}\right], \\
\sigma_{DR}^2 &= \left(2 + e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\right)\sigma_\varepsilon^2, \\
\sigma_{M,\hat{p}}^2 &= \left(\sigma_{Y_w}^2 - \frac{\sigma_{Y_w,Z}^2}{\sigma_Z^2}\right)\left(\frac{1 + 4M + (e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2})(1+2M)}{2M}\right) \\
&- \frac{1}{E\left[p'(Z)\right]}\left(\sigma_{Y_w}^2 - \sigma_\varepsilon^2 - \frac{\sigma_{Y_w,Z}^2}{\sigma_Z^2}\right),
\end{aligned}$$

*with the choice of* $w = 0, 1$ *being arbitrary.*

Proposition 1 allows us to establish the following corollary on the behaviour of the estimators with respect to the correlation.

**Corollary 2** *Under Assumptions 1–3, if all elements in $\gamma$ and $\beta_w$ have equal signs, then an increase in the correlation will lead to an increase in $\sigma^2_{IPW,\hat{p}}$, an increase in $\sigma^2_{DR}$ and an increase or a decrease in $\sigma^2_{M,\hat{p}}$.*

The behaviour of the IPW estimator is not straightforward and therefore a proof is provided in the appendix. The effect of correlation on the variance of the DR estimator is obvious by inspection, due to the quadratic form of $\sigma^2_Z$. The behaviour of the matching estimator is easily shown by a numerical counter example, e.g., increasing the correlation when $\beta = (0.2, 2)$ and $\gamma = (0.3, 0.3)$ gives a smaller $\sigma^2_{\mathrm{M},\hat{p}}$, while increasing the correlation when $\beta = (0.2, 2)$ and $\gamma = (0.8, 0.8)$ gives a larger $\sigma^2_{\mathrm{M},\hat{p}}$.

Some of our findings deserve special mention. The results for the IPW estimator are consistent with previous result for the IPW estimator using the true propensity score (Pingel and Waernbaum, 2014). The DR estimator exhibits the same behaviour as the IPW estimator with regard to direction, but we also observe that $\sigma^2_{\mathrm{DR}}$ is only affected by the correlation through the parameter vector $\gamma$ in the treatment assignment. Finally, the results for the matching estimator are similar to those for the matching estimator using the true propensity score shown in Pingel and Waernbaum (2014) in that $\sigma^2_{\mathrm{M},\hat{p}}$ may increase or decrease.

# 6 Numerical results and simulation studies

Because the analytic results are restricted to the assumed data generating process (DGP), this section is devoted to numerical studies of the estimators where we relax some of the previous assumptions. This includes studying the effect of correlation under different true and assumed causal structures, for instance when not all covariates are confounders or when we fail to include a confounder in the estimation model. Furthermore, the asymptotic results are extended through simulation to include the finite sample properties of the estimators. The asymptotic behaviour of the estimators is evaluated using numerical methods, while the finite sample properties are studied for the sample size of $N = 1000$, using simulations with 5000 replicates.

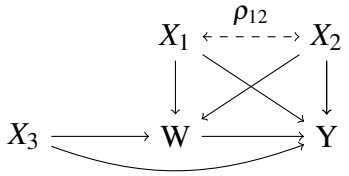


**Figure 1:** Causal diagram 1

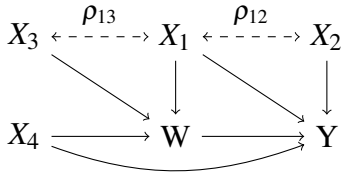


**Figure 2:** Causal diagram 2

### 6.1 Case 1: Non-constant causal effect

The aim of *Case 1* is to analyse the effect of correlation on the variances of the estimators when varying the magnitude of the parameters in the outcome models and treatment assignment model. This manipulation allows for investigation of how the strength of the confounding towards the outcome and treatment assignment is related to the effect of the correlation. Moreover, it enables us to study the effect of correlation when deviating from the assumption of constant causal effect in Assumption 3.

Assume that Assumptions 1–3(*i–ii*) hold. The causal structure of this study design consists of three standardised variables, $X_1$, $X_2$, $X_3$, and is displayed in Figure 1. We let the correlations take the values $\rho_{13} = 0$, $\rho_{23} = 0$, and $\rho_{12} = 0, 0.1, \ldots, 0.9, 0.95, 0.97, 0.99$. The parameters $\beta_{11}$, $\beta_{12}$, $\beta_{01}$, $\beta_{02}$ take values from the set $\{0.5, 2\}$ and the parameters $\gamma_1$ and $\gamma_2$ take values from the set $\{0.2, 0.4, 0.8, 1.2\}$. We include $X_3$ as noise to ensure that $\text{Cov}(X, Y_w | p(Z))$ will not approach zero when $\rho_{12}$ tends to one. In addition, $\beta_{13} = \beta_{03} = 1$ and $\gamma_3 = 0.3$. Thus, all covariates are confounders. The intercepts of the outcome models are set to $\alpha_1 = 5$ and $\alpha_0 = 0$. Finally, we let $\sigma^2_\varepsilon = 1$.

The results of the simulation study are displayed in Figure 3, which presents the asymptotic standard errors and the finite sample standard errors of the estimators.

First, we observe that under non-constant causal effect, a change in $\rho_{12}$ may in fact decrease the asymptotic variances of the IPW and DR estimator. This behaviour is seen for the case when $\beta_{11} = 2$, $\beta_{12} = 0.5$, $\beta_{01} = 0.5$, $\beta_{02} = 1$, $\gamma_1 = 0.2$, and $\gamma_2 = 0.4$. Second, the variance of the IPW estimator, in comparison with the DR and matching estimators, is more extremely affected by a change in $\rho_{12}$. The DR and matching estimators exhibit a similar degree of sensitivity to a change in $\rho_{12}$, albeit in different directions. By visual inspection we conjecture that $\partial \sigma^2_{\text{M},\hat{p}} / \partial \rho_{12} < \partial \sigma^2_{\text{IPW},\hat{p}} / \partial \rho_{12}$, which was also suggested in Pingel and Waernbaum (2014) through a series of examples. Third, the asymptotic and

finite sample variances overlap in most cases, but when too much density is in the tails of the propensity score distribution, the asymptotic distributions of the estimators fail to approximate the finite sample distributions. This is in accord with the findings of Pingel and Waernbaum (2014), which is related to Kahn and Tamer's 2010 results. A conclusion is that an increase in $\rho_{12}$ may increase an already strong treatment assignment making the asymptotic results no longer valid. As a remark, given that the asymptotic and finite sample variances do not overlap, the finite sample variances of the IPW and matching estimators tend to be smaller than the asymptotic variances, whereas for the DR estimator, the asymptotic variance underestimates the finite sample variance.

Similar to Pingel and Waernbaum (2014), although some patterns can be observed in how $\beta$ and $\gamma$ determine how $\rho_{12}$ affects the variances concerning the direction and magnitude, we conclude that predictions of the effect on the variances are difficult to make in practice.

### 6.2 Case 2: Non-normal covariates

*Case 2* concerns whether deviations from normality influence the effect $\rho_{12}$ has on the estimators. Here, we consider the DGP in *Case 1*, but only allow for a constant causal effect. Further, let $X_1$ and $X_2$ follow a $U(-1.5, 1.5)$ distribution.

The following results can be seen in Figure 4. First, the curvatures of the standard errors as functions of $\rho_{12}$ are smaller than those in Figure 3. Second, the increase in the standard error of the IPW estimator is not as extreme as that in Figure 3. An explanation is that correlated variables with a finite support, such as uniformly distributed covariates, can be bounded away from extreme values of the propensity score. Third, we observe that the overall pattern for the direction of the effect is similar to *Case 1*.

Not shown in this paper but available from the authors upon request are results when $X_1$ follows a Poisson distribution and $X_2$ follows a gamma distribution. Again, because the support of $X_1$ and $X_2$ is infinite, a change in $\rho_{12}$ may yield a large increase in the variance of the IPW estimator. However, when the propensity score distribution is well-behaved, the overall effect of $\rho_{12}$ on the estimators is smaller compared to when $X_1$ and $X_2$ are normally distributed.

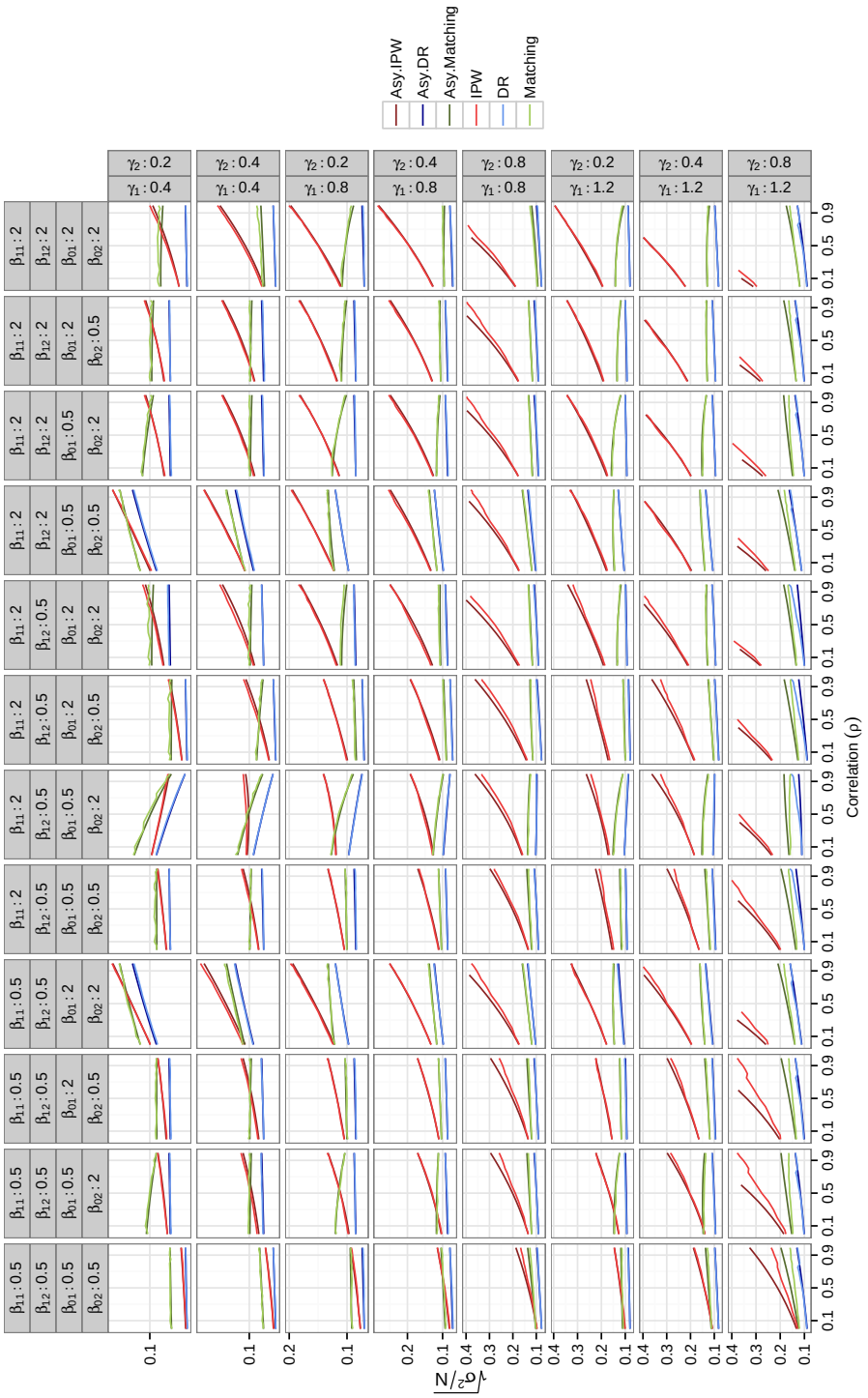


**Figure 3:** Effect of $\rho_{12}$ on asymptotic and finite sample standard errors in Case 1.

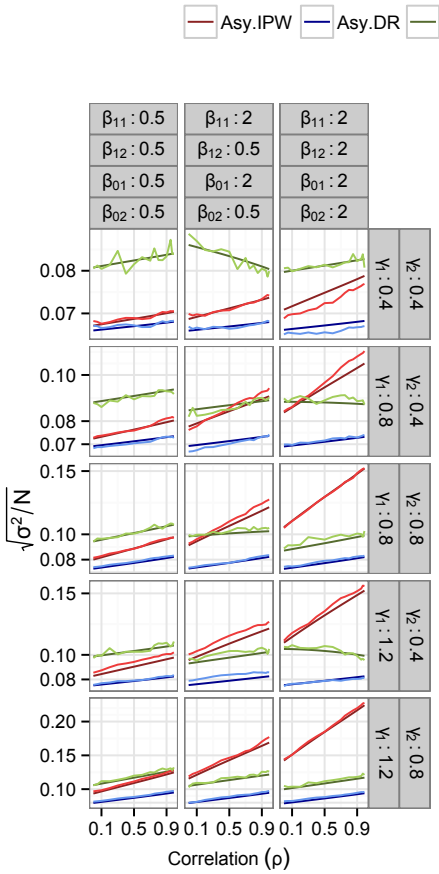


**Figure 4:** Effect of $\rho_{12}$ on asymptotic and finite sample standard errors in Case 2.

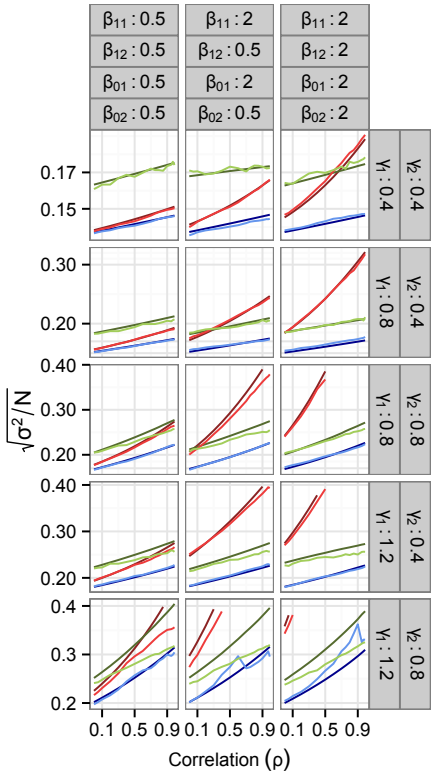


**Figure 5:** Effect of $\rho_{12}$ on asymptotic and finite sample standard errors in Case 3.

### 6.3 Case 3: Non-constant error variance

In *Case 3* we study whether non-constant variance of the error term in the outcome models could change the results in Proposition 1. Consider the DGP in *Case 1*, but only allow for a constant causal effect. Further, let the error variances in the outcome models be generated according to

$$\varepsilon = \varepsilon^*[0.5 + 3p(Z)], \ \varepsilon^* \sim N(0,1).$$

Figure 5 shows that for $\beta_{11} = 2$, $\beta_{12} = 0.5$, $\beta_{01} = 2$, $\beta_{02} = 0.5$, $\gamma_1 = 0.4$ and $\gamma_2 = 0.4$ we have increasing variance of the matching estimator as a function of $\rho_{12}$ instead of decreasing variance, which was the result in *Case 1*. Thus, correlation between covariates may affect the estimators through the error terms of the outcome models. We also see

that when the treatment assignment is strong, the DR estimator displays some erratic behaviour for the finite sample variance and that the asymptotic variance of the matching estimator overestimates the finite sample variance.

### 6.4 Case 4: Omitting a confounder

*Case 4* studies the effect of $\rho_{12}$ on the behaviour of the estimators when a correlated confounder is omitted. This setting is different from *Case 1–3* in that the unconfoundedness assumption, Assumption $1(i)$, is not fulfilled. *Case 4* is motivated by our previous findings in which we discovered that an increase in $\rho_{12}$ in some cases yielded a very large increase in the variance of the IPW estimator. Because the increase is due to the correlation, omitting a confounder will remove the effect that $\rho_{12}$ has on the variance. The omission of a confounder results in biased estimates; however, a large $\rho_{12}$ implies that two confounders share much of the same information, resulting in a trade-off between bias and variance. We therefore study how the correlation affects the bias and mean squared error (MSE) of the estimators when omitting a confounder.

Consider the DGP in *Case 1* and let $\hat{\tau}(X_1, X_2, X_3)$ denote an estimator of $\tau$ using $X_1$, $X_2$ and $X_3$ (i.e. all confounders) and let $\hat{\tau}(X_1, X_3)$ denote an estimator of the average causal effect using $X_1$ and $X_3$ only. To evaluate the estimators we estimate Relative MSE $=$ $\text{MSE}[\hat{\tau}(X_1, X_3)]/\text{MSE}[\hat{\tau}(X_1, X_2, X_3)] = E([\hat{\tau}(X_1, X_3) - \tau]^2)/E([\hat{\tau}(X_1, X_2, X_3) - \tau]^2)$, and $\text{Bias} = E[\hat{\tau}(X_1, X_3) - \tau]$. The results for the relative MSE and bias are depicted in Figure 6 and Figure 7, respectively.

The conclusion for the DR and matching estimator when evaluating their MSE is clear, that is all confounders should, for all models in the simulations, be included in the propensity score model, no matter the correlation. As for the IPW estimator, we observe that in some cases, when the treatment assignment is strong, it is beneficial in terms of MSE to omit a confounder from the propensity score model. However, in most cases all covariates should be included for the IPW estimator as well. We also observe that the MSE of the studied IPW estimator is less negatively affected by an omitted confounder than the other estimators, probably because of the large decrease in the variance. We also included an OLS estimator as a comparison, which proved to be the estimator most

negatively affected in terms of MSE when omitting a confounder.

Regarding bias, we observe a similar pattern for all estimators, i.e. an increased $\rho_{12}$ leads to a non-linear decrease in bias. The bias for the matching estimator seems to be slightly larger for some specifications of the treatment assignment model compared with the other estimators.

### 6.5 Case 5: Inclusion of irrelevant covariates

*Case 5* is an extension of the simulations studies by Brookhart et al. (2006) and Millimet and Tchernis (2009), which, in turn, are related to the theoretical results in Rubin and Thomas (1996). These studies concern the efficiency of propensity score estimators when covariates related to only the outcome or the treatment assignment are included in the propensity score model. To summarise their findings, it is beneficial in terms of MSE to include covariates related to outcome, but not the treatment assignment. Covariates that are only related to the treatment should not be included in the model. The purpose of this paper is to study if and how these results are affected when correlation is included in the analysis.

We use a simulation design that resembles those in Brookhart et al. (2006) and Millimet and Tchernis (2009). Let $X_1$, $X_2$, $X_3$ and $X_4$ be distributed according to a standard uniform distribution and consider the causal structure in Figure 2. The covariates $X_1$ and $X_4$ are confounders and should always be included in the propensity score for the estimators to be consistent. Although $X_2$ is only related to the outcome, it should, according to the findings in the aforementioned studies, be included in the propensity score in order to gain efficiency. However, the covariate $X_3$ that is only related to the treatment assignment should not be included in the propensity score since that would incur some efficiency loss.

As seen in Figure 2, *Case 5* includes two correlations and we study either the effect of correlation between a confounder, $X_1$, and the covariate related to the outcome only, $X_2$, or the effect of the correlation between a confounder, $X_1$, and the covariate related to the treatment only, $X_3$. Let the correlation matrix in the DGP be either

$$(\rho_{st}) = \begin{pmatrix} 1 & \rho_{12} & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{pmatrix} \text{ or } (\rho_{st}) = \begin{pmatrix} 1 & 0 & \rho_{13} & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{pmatrix},$$

where $\rho_{12} = \rho_{13} = 0, 0.1, \ldots, 0.9, 0.95, 0.99$. The treatment assignment model is given by

$$\text{logit}(p) = 0.25[(-3 + 6X_1) + (-3 + 6X_3) + (-3 + 6X_4)].$$

Compared with the design in Millimet and Tchernis (2009), the strength of the treatment assignment is slightly reduced, so that $\min(p) \approx 0.1$, $\max(p) \approx 0.9$, instead of $\min(p) \approx 0.05$, $\max(p) \approx 0.95$. We study both a constant and a non-constant causal effect. The potential outcome models yielding a constant causal effect are

$$Y_1 = 0.87 + 2.47X_1 + 2.47X_2 + 2.47X_4 + \varepsilon_1,$$
$$Y_0 = 2.54 + 2.47X_1 + 2.47X_2 + 2.47X_4 + \varepsilon_0,$$

while the potential outcome models resulting in a non-constant causal effect are

$$Y_1 = 0.87 + 0.51X_1 + 0.51X_2 + 0.51X_4 + \varepsilon_1$$
$$Y_0 = 2.54 + 2.47X_1 + 2.47X_2 + 2.47X_4 + \varepsilon_0.$$

In the outcome models, $\varepsilon_1, \varepsilon_0 \sim N(0, 0.25)$. For each generated data set, we estimate the five propensity score models including the covariates (1) $\{X_1\}$, (2) $\{X_1, X_2\}$, (3) $\{X_1, X_2, X_3\}$, (4) $\{X_1, X_3\}$ or (5) $\{X_2, X_3\}$. Because $X_4$ is always included, we omit $X_4$ in the notation. For the DR estimator, the same covariates are included in the regression parts as in the propensity score model. The estimators are evaluated by estimating $\text{MSE} = E[(\hat{\tau} - \tau)^2]$.

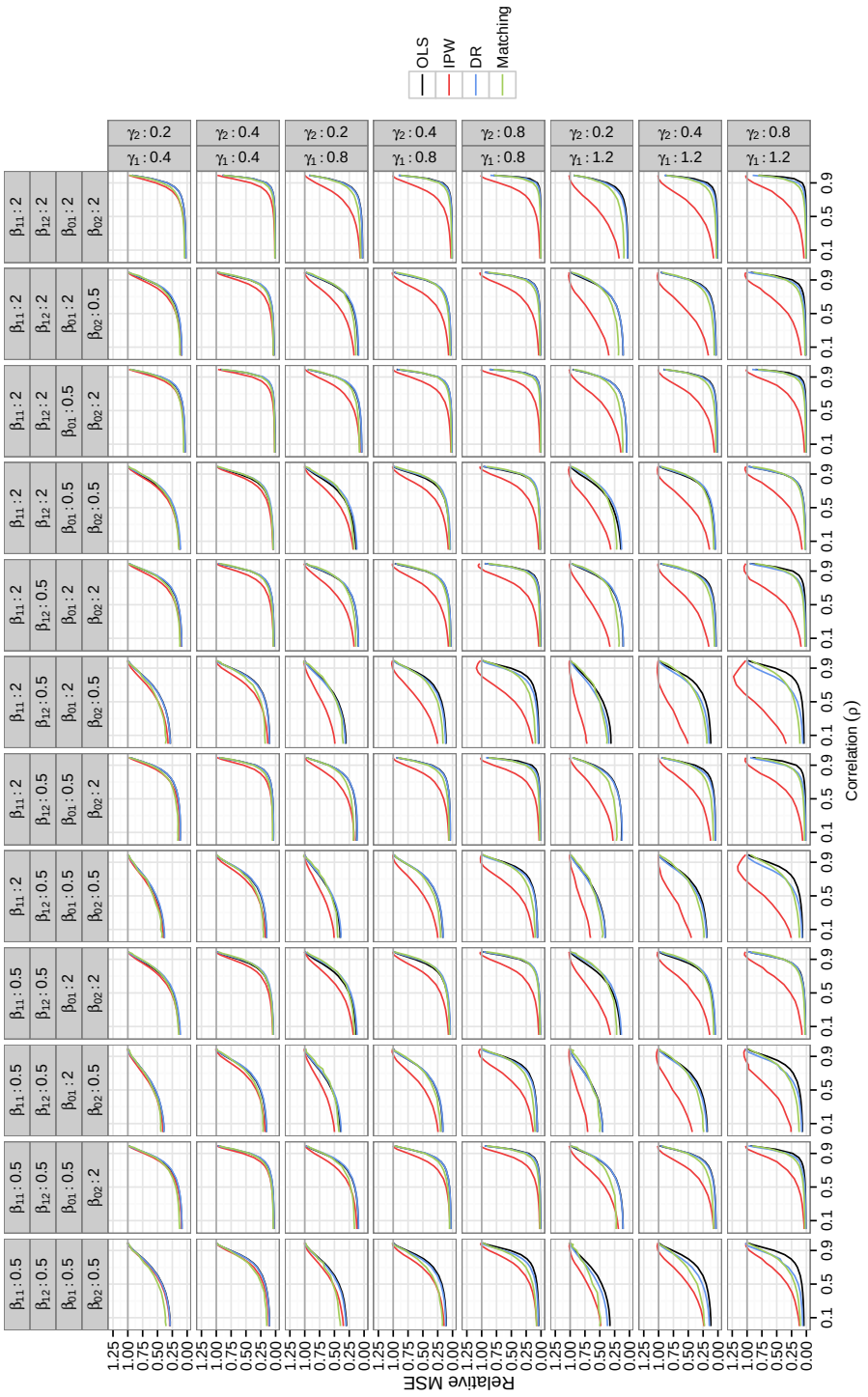


**Figure 6:** Effect of $\rho_{12}$ on Relative MSE in Case 4.

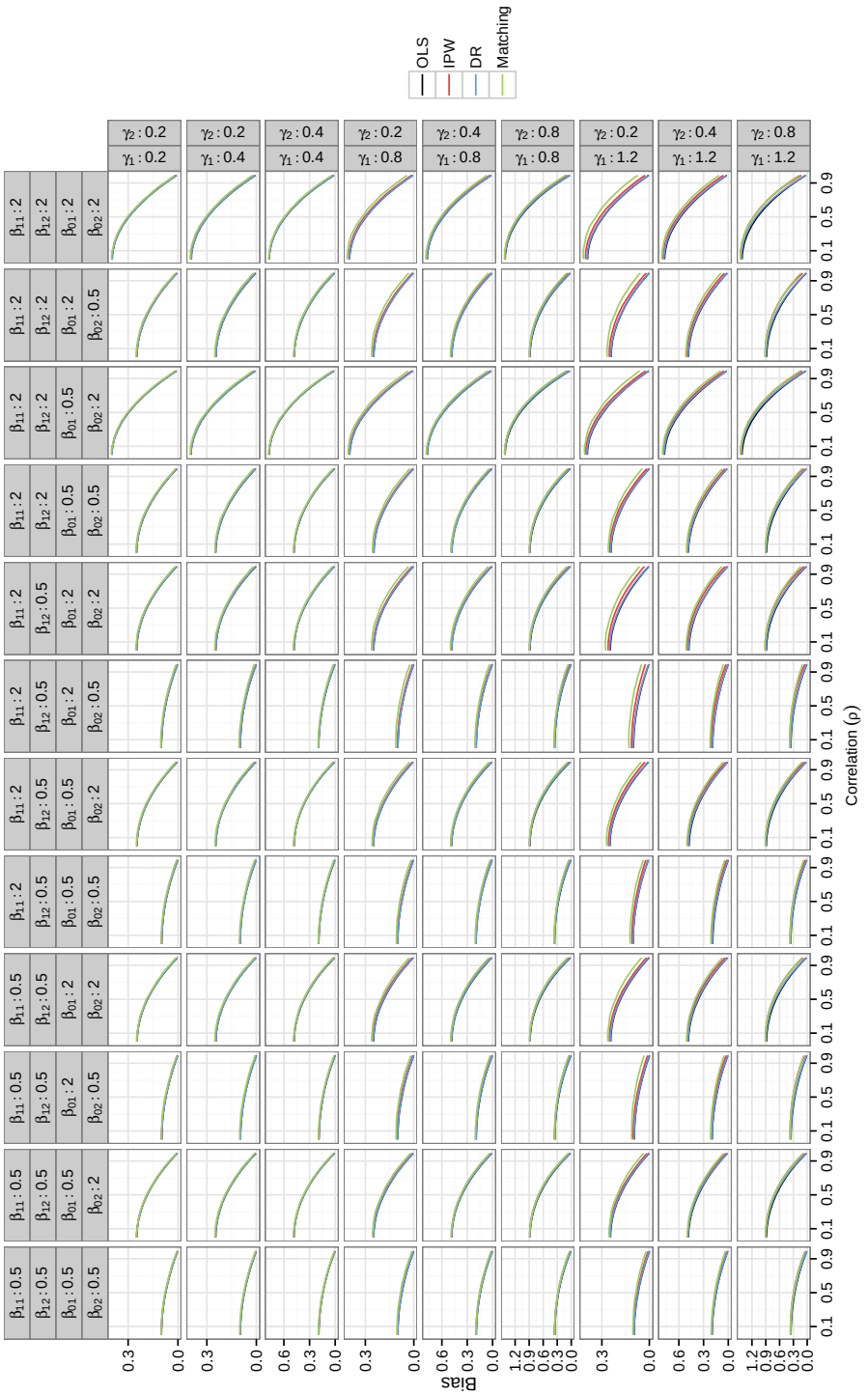


**Figure 7:** Effect of $\rho_{12}$ on bias in Case 4.

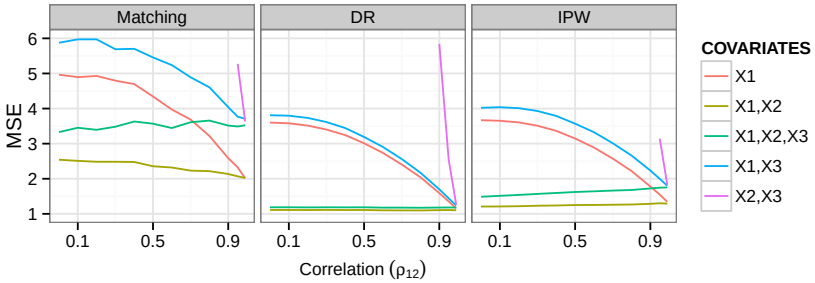


**Figure 8:** Effect of $\rho_{12}$ on MSE when having a constant causal effect.

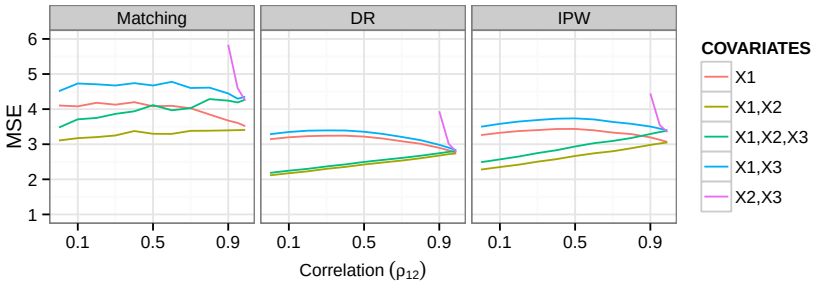


**Figure 9:** Effect of $\rho_{12}$ on MSE when having a non-constant causal effect.

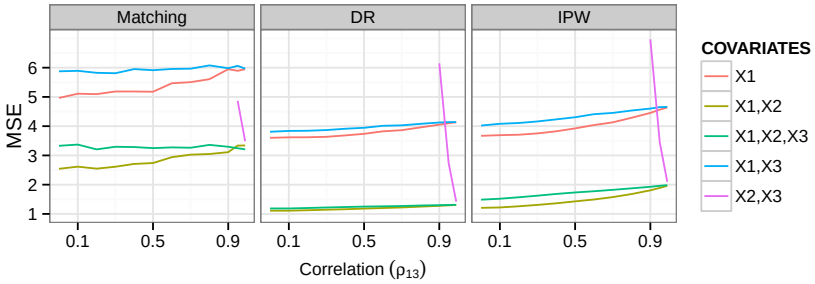


**Figure 10:** Effect of $\rho_{13}$ on MSE when having a constant causal effect.

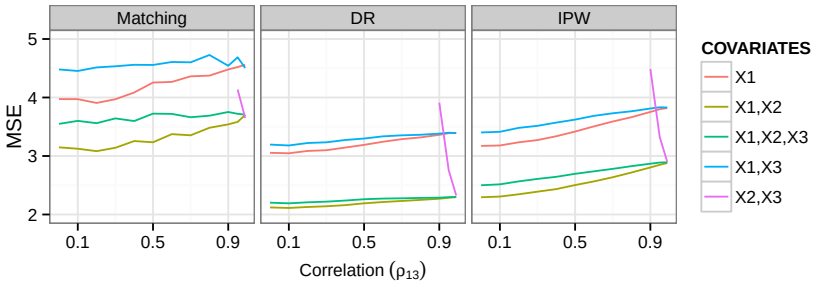


**Figure 11:** Effect of $\rho_{13}$ on MSE when having a non-constant causal effect.

The results are displayed in Figures 8–11, which show that the estimators are most efficient when $X_1, X_2$ are included in the propensity score model. The matching estimator is unique in that it, for a propensity score model that includes $X_1, X_2$, may become more efficient when $\rho_{12}$ increases, as seen for the case with constant causal effect. Moreover, we observe for the matching and IPW estimators that a large $\rho_{12}$ can lead to that estimators with $X_1$ only become more efficient than estimators including $X_1, X_2, X_3$. This behaviour is not seen for the DR estimator. Instead, we see a small cost in terms of efficiency loss by including an irrelevant covariate related to treatment in the DR estimator for all correlations. This can be explained by the fact that the DR estimator has correctly specified outcome models. Thus, a positive trade-off between bias and variance as the correlation increases is absent.

As for the effect of $\rho_{13}$ on the MSE of the estimators, the behaviour is the same for the constant and non-constant causal effect, i.e. an increase in $\rho_{13}$ results in most of the estimators becoming less efficient. Particularly noteworthy is that the estimator that includes $X_2, X_3$ (i.e. omitting one of the confounders) is actually more efficient for large $\rho_{13}$ than estimators that include $X_1$ or $X_1, X_3$.

# 7 Conclusions

In this study we examine how correlation affects commonly used propensity score-based estimators of the average causal effect. This examination involve formalising how correlation between observed covariates influences the variance of the estimators for the average causal effect.

We then proceed to show under specific model assumptions (such as constant causal effect and normally distributed covariates) that an increase in the correlation leads to an increase in the asymptotic variance of the IPW and DR estimator if the model parameters in the outcome and treatment models share the same sign. The variance of the propensity score matching estimator, however, can both decrease and increase.

To extend the analytic results we perform numerical and finite-sample investigations for a wide range of models. Here, we see results in different directions. For instance, if

the causal effect is heterogeneous, the variance of the IPW estimator can both increase and decrease, although for the models under study the overall impression is that the IPW estimator shows the most instability with respect to the change in correlation.

We also study the bias-variance trade-off by omitting a correlated confounder in the propensity score model. For some cases in this scenario we observe a decrease in the MSE, but only when the propensity scores are extreme, i.e. close to zero or one.

Finally, we see that including an irrelevant but correlated covariate affects the efficiency of the estimators, something that could influence covariate selection. These findings contributes to the results by Brookhart et al. (2006) and Millimet and Tchernis (2009).

Although it is difficult to rationalise guidelines based on the results of this study, we were able to demonstrate that the correlation between covariates could prove to be an important aspect to consider when modelling the propensity score.

# References


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

Abadie, A. and G. W. Imbens (2012). Matching on the estimated propensity score. Harvard University and National Bureau of Economic Research. http://www.hks.harvard.edu/fs/aabadie/pscore.pdf. Retrieved 2014-08-01.

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

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement 7*, 249–253.

Connors, A. F., T. Speroff, N. V. Dawson, C. Thomas, F. E. Harrell, D. Wagner, N. Desbiens, L. Goldman, A. W. Wu, R. M. Califf, W. J. Fulkerson, H. Vidaillet, S. Broste, P. Bellamy, J. Lynn, and W. A. Knaus (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *Journal of the American Medical Association 276*, 889–897.

de Luna, X., I. Waernbaum, and T. S. Richardson (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika 98*, 861–875.

Hahn, J. (2004). Functional restriction and efficiency in causal inference. *Review of Economics and Statistics 86*, 73–76.

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

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

Kahn, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*, 2021–2042.

Lee, K. W. (1990). Bootstrapping logistic regression models with random regressors. *Communications in Statistics - Theory and Methods 19*, 2527–2539.

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.

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. English translation by D. Dabrowska and T. Speed in *Statistical Science 5*, 465–472, 1990.

Pingel, R. (2014). Some approximations of the logistic distribution with application to the covariance matrix of logistic regression. *Statistics & Probability Letters 85*, 63–68.

Pingel, R. and I. Waernbaum (2014). Effects of correlated covariates on the asymptotic efficiency of matching and inverse probability weighting estimators for causal inference. *Statistics: A Journal of Theoretical and Applied Statistics*. Epub ahead of print. doi: 10.1080/02331888.2014.925899.

Robins, J. M., A. Rotnitsky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*, 846–866.

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.

Rubin, D. B. (1980). Randomization analysis of experimental data: the Fisher randomization test comment. *Journal of the American Statistical Association 75*, 591–593.

Rubin, D. B. and N. Thomas (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics 52*, 249–264.

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

Stürmer, T., M. Joshi, R. J. GLynn, J. Avorn, K. J. Rothman, and S. Schneeweiss (2006). A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with coventional multivariable methods. *Journal of Clinical Epidemiology 59*, 437–447.

Vansteelandt, S., M. Bekaert, and G. Claeskens (2010). On model selection and model misspecification in causal inference. *Statistical Methods in Medical Research 21*, 7–30.

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

White, H. and X. Lu (2011). Causal diagrams for treatment effect estimation with application to efficient covariate selection. *The Review of Economics and Statistics 93*, 1453–1459.

# Appendix

Assume that Assumptions 1–3 hold. We omit, for brevity, the unit vector in $X$ when we describe the multivariate normal (MVN) distribution. Following the properties of the MVN distribution (e.g., Mardia et al., 1979), $X$, $Z$ and a potential outcome, $Y_w$, have a $k+2$ dimensional MVN distribution with covariance matrix

$$\Lambda = \begin{pmatrix} \Sigma & \sigma_{X,Y_w} & \sigma_{X,Z} \\ \sigma'_{X,Y_w} & \sigma^2_{Y_w} & \sigma_{Y_w,Z} \\ \sigma'_{X,Z} & \sigma_{Y_w,Z} & \sigma^2_Z \end{pmatrix} = \begin{pmatrix} \Sigma & \Sigma\beta_w & \Sigma\gamma \\ \beta'_w\Sigma & \beta'_w\Sigma\beta_w + \sigma^2_\varepsilon & \beta'_w\Sigma\gamma \\ \gamma'\Sigma & \beta'_w\Sigma\gamma & \gamma'\Sigma\gamma \end{pmatrix},$$

where $\Sigma$ denotes the $k \times k$ covariance matrix of $X$ with variances $\sigma^2_{X_s}$ and covariances $\sigma_{X_s,X_t}$. Also, $\sigma'_{X,Y_w} = (\sigma_{X_1,Y_w}, \ldots, \sigma_{X_k,Y_w})$ and $\sigma'_{X,Z} = (\sigma_{X_1,Z}, \ldots, \sigma_{X_k,Z})$. The MVN distribution has the conditional moments $E(Y_w|Z) = \mu_{Y_w} + \sigma_{Y_w,Z}(Z-\mu_Z)/\sigma^2_Z$, $E(X|Z) = \sigma_{X,Z}(Z-\mu_Z)/\sigma^2_Z$ and $\mathrm{Cov}(X, Y_w|Z) = \sigma_{X,Y_w} - \sigma_{X,Z}\sigma_{Y_w,Z}/\sigma^2_Z$.

Theorem 1 in Pingel (2014) states that if $X$ is MVN (and including the unit vector), the inverse of information matrix in Equation (1) can be written

$$I^{-1} = \begin{pmatrix} I^{11} & & I^{12} \\ I^{21} & \dfrac{\Sigma^{-1}}{E[p'(Z)]} - \dfrac{\gamma\gamma'}{E[p'(Z)]\sigma^2_Z} + \dfrac{\gamma\gamma' E[p'(Z)]}{E[p'(Z)]E[p'(Z)Z^2] - (E[p'(Z)Z])^2} \end{pmatrix}.$$

Observe that because $p(Z) = e^Z/(1+e^Z)$, we have that $p'(Z) = e^Z/(1+e^Z)^2$.

Recall the large sample variances in Pingel and Waernbaum (2014) and reformulate them without assuming that $\mu_Z = 0$,

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

**Proof of Proposition 1.** To find $a'I^{-1}a$, we rewrite Equation (5):

$$\begin{aligned}
a =& E\left[\left(\frac{Y_1-\mu_1}{p(Z)}+\frac{Y_0-\mu_0}{1-p(Z)}\right)\frac{e^Z}{(1+e^Z)^2}X\right]\\
=& E\left[E\left(\frac{XY_1}{1+e^Z}+\frac{e^ZXY_0}{1+e^Z}-\mu_{Y_1}\frac{X}{1+e^Z}-\mu_{Y_0}\frac{e^ZX}{1+e^Z}|Z\right)\right]\\
=& E\left[(1-p(Z))\left(\mathrm{Cov}(X,Y_1|Z)+E(Y_1|Z)E(X|Z)-\mu_{Y_1}E(X|Z)\right)\right.\\
&+ p(Z)\left(\mathrm{Cov}(X,Y_0|Z)+E(Y_0|Z)E(X|Z)-\mu_{Y_0}E(X|Z)\right)]\\
=& E\left[(1-p(Z))\left(\mathrm{Cov}(X,Y_1|Z)+\frac{\sigma_{Y_1,Z}}{\sigma_Z^2}(Z-\mu_Z)E(X|Z)\right)\right.\\
&\left.+ p(Z)\left(\mathrm{Cov}(X,Y_0|Z)+\frac{\sigma_{Y_0,Z}}{\sigma_Z^2}(Z-\mu_Z)E(X|Z)\right)\right]=\sigma_{X,Y_w}.
\end{aligned}$$

The last equality is due to the constant causal effect. From the constant causal effect it follows, after substituting the first element in $X$ with unity representing the intercept, that the first element in $a$ is 0. After some basic algebra

$$a'I^{-1}a=\frac{\sigma_{Y_w}^2-\sigma_\varepsilon^2}{E\left[p'(Z)\right]}-\frac{\sigma_{Y_w,Z}^2}{E\left[p'(Z)\right]\sigma_Z^2}+\frac{\sigma_{Y_w,Z}^2E\left[p'(Z)\right]}{E\left[p'(Z)\right]E\left[p'(Z)Z^2\right]-\left(E\left[p'(Z)Z\right]\right)^2}.$$

To find $b$, let $Y_w^*\equiv E(Y_w|X)-\mu_w$, where $Y_w^*\sim N\left(0,\sigma_{Y_w}^2-\sigma_{\varepsilon_w}^2\right)$, $w=0,1$. Equation (8) is written $b=E\left[e^Z(Y_0^*)^2+2Y_0^*Y_1^*+e^{-Z}(Y_1^*)^2\right]$. Since $E(Y_w|X)=\alpha_w+\beta_w'X$, $E(Y_0^*Y_1^*)=\sigma_{Y_0,Y_1}^2=\beta_1'\Sigma\beta_0$. Next, we use that $E\left(Y_we^Z\right)=e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w}+\sigma_{Y_w,Z}\right)$ and $E\left(Y_we^{-Z}\right)=e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\mu_{Y_w}-\sigma_{Y_w,Z}\right)$, as shown in Pingel and Waernbaum (2014), and $b$ can be formulated

$$b=e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_0}^2-\sigma_{\varepsilon_0}^2+\sigma_{Y_0,Z}^2\right)+2\sigma_{Y_0,Y_1}+e^{-\mu_Z+\frac{1}{2}\sigma_Z^2}\left(\sigma_{Y_1}^2-\sigma_{\varepsilon_1}^2+\sigma_{Y_1,Z}^2\right).$$

To find $c'I^{-1}c$, observe that due to the constant causal effect $\mathrm{Cov}(X,Y_1|p(Z))=\mathrm{Cov}(X,Y_0|p(Z))=\mathrm{Cov}(X,Y_w|p(X))$. Hence, after simplification of Equation (12), $c=E[\mathrm{Cov}(X,Y_w|p(X)]$. Since Assumption 3 implies a constant variance,

$$c=\sigma_{X,Y_w}-\sigma_{X,Z}\sigma_{Y_w,Z}/\sigma_Z^2.$$

Including the unit vector in $X$, the first element in $c$ is equal to 0. After some algebra

$$c'I^{-1}c=\left(\sigma_{Y_w}^2-\sigma_\varepsilon^2-\sigma_{Y_w,Z}^2/\sigma_Z^2\right)/E\left[p'(Z)\right].$$

After subtraction of the adjustment terms from $\sigma^2_{\text{IPW}}$ and $\sigma^2_{\text{M}}$ and simplifying we arrive at Proposition 1. ■

**Proof of Corollary 1 (IPW estimator).** Define the integrals

$$A \equiv \int_{-\infty}^{\infty} \frac{e^z}{(1+e^z)^2} e^{-\frac{(z-\mu_Z)^2}{2\sigma_Z^2}} \, \mathrm{d}z, \;\; B \equiv \int_{-\infty}^{\infty} \frac{ze^z}{(1+e^z)^2} e^{-\frac{(z-\mu_Z)^2}{2\sigma_Z^2}} \, \mathrm{d}z,$$
$$C \equiv \int_{-\infty}^{\infty} z^2 \frac{e^z}{(1+e^z)^2} e^{-\frac{(z-\mu_Z)^2}{2\sigma_Z^2}} \, \mathrm{d}z.$$

Rewriting $\sigma^2_{\text{IPW},\hat{p}}$ we have

$$\begin{aligned}\sigma^2_{\text{IPW},\hat{p}} =& (\sigma^2_{Y_w} - \sigma^2_\varepsilon)\left(2 + e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2} - \frac{\sqrt{2\pi}\sigma_Z}{A}\right) \\ &+ \sigma^2_{Y_w,Z}\left(e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2} + \frac{\sqrt{2\pi}}{A\sigma_Z} + \frac{A\sqrt{2\pi}\sigma_Z}{B^2-AC}\right) \\ &+ \left(2 + e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2}\right)\sigma^2_\varepsilon.\end{aligned}$$

After evaluation of this expression we see that without loss of generality we can assume that $\mu_Z = 0$, since for $\mu_Z = 0$ the exponential parts $e^{-\mu_Z+\frac{1}{2}\sigma_Z^2} + e^{\mu_Z+\frac{1}{2}\sigma_Z^2}$, that dominate the increase, have the smallest growth relative to the other terms. Rewriting we get

$$\begin{aligned}\sigma^2_{\text{IPW},\hat{p}} =& (\sigma^2_{Y_w} - \sigma^2_\varepsilon)\underbrace{\left(2 + 2e^{\frac{1}{2}\sigma_Z^2} - \frac{\sqrt{2\pi}\sigma_Z}{A}\right)}_{\Psi} \\ &+ \sigma^2_{Y_w,Z}\underbrace{\left(2e^{\frac{1}{2}\sigma_Z^2} + \frac{\sqrt{2\pi}}{A\sigma_Z} + \frac{A\sqrt{2\pi}\sigma_Z}{B^2-AC}\right)}_{\Omega} + \left(2 + 2e^{\frac{1}{2}\sigma_Z^2}\right)\sigma^2_\varepsilon.\end{aligned}$$

First, we note that $\sigma^2_Z$, $\sigma^2_{Y_w}$ and $\sigma^2_{Y_w,Z}$ are increasing functions of $\rho$ when all elements in $\gamma$ and $\beta_w$ have equal signs. We also note that $\sigma^2_\varepsilon$ is a constant independent of $\rho$. Since we aim at investigating the behaviour of $\sigma^2_{\text{IPW},\hat{p}}$ as $\rho$ increases it is sufficient to estimate the contribution of the components inside the parentheses when $\sigma_Z$ increases, given that $\sigma^2_{Y_w} > \sigma^2_\varepsilon$. There are no closed form expressions for $A$, $B$ and $C$, instead we use definite integrals with large limits which serves as approximations. We write $A(\sigma_Z)$, $B(\sigma_Z)$, $C(\sigma_Z)$ to emphasise the integrals as functions

of $\sigma_Z$.

Evaluating the first parenthesis, $\Psi$, we see that for $\sigma_Z > 0$, $\Psi$ is an increasing function of $\sigma_Z$ (Figure A.1). Similarly, we investigate the second parenthesis, $\Omega$, which increases with $\sigma_Z$ (Figure A.2). Since $\sigma_\varepsilon > 0$ and a constant we see that the third term is an increasing function of $\sigma_Z$. Hence, we conclude that $\sigma^2_{\text{IPW},\hat{p}}$ increases with $\rho$.

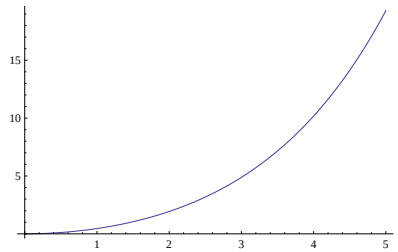


**Figure A.1:** $\Psi(\sigma_Z)$

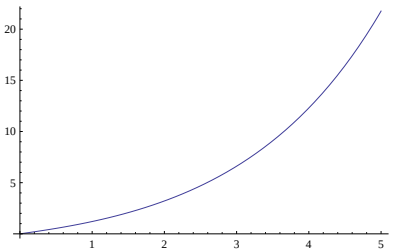


**Figure A.2:** $\Omega(\sigma_Z)$

■

## Publication series published by IFAU – latest issues

### Rapporter/Reports

**2014:20** Sibbmark Kristina ”Arbetsmarknadspolitisk översikt 2013”

**2014:21** Nordlund Madelene and Mattias Strandh ”Selektivitet och jobbchanser bland arbetslösa”

**2014:22** Angelov Nikolay and Marcus Eliason ”Vilka arbetssökande kodas som funktionshindrade av Arbetsförmedlingen?”

**2014:23** Angelov Nikolay and Marcus Eliason ”Friställd och funktionsnedsatt”

**2014:24** Angelov Nikolay and Marcus Eliason ”Lönebidrag och skyddat arbete: en utvärdering av särskilda insatser för sökande med funktionshinder”

**2014:25** Holmlund Helena, Josefin Häggblom, Erica Lindahl, Sara Martinson, Anna Sjögren, Ulrika Vikman and Björn Öckert ”Decentralisering, skolval och friskolor: resultat och likvärdighet i svensk skola”

**2014:26** Lindgren Karl-Oskar, Sven Oskarsson and Christopher Dawes ”Kan politisk ojämlikhet utbildas bort?”

### Working papers

**2014:20** Johansson Per, Lisa Laun and Mårten Palme “Pathways to retirement and the role of financial incentives in Sweden”

**2014:21** Andersson Elvira, Petter Lundborg and Johan Vikström “Income receipt and mortality – evidence from Swedish public sector employees”

**2014:22** Felfe Christina and Rafael Lalive “Does early child care help or hurt children’s development?”

**2014:23** Nordlund Madelene and Mattias Strandh “The relation between economic and non-economic incentives to work and employment chances among the unemployed”

**2014:24** Mellander Erik “Transparency of human resource policy”

**2014:25** Angelov Nikolay and Marcus Eliason “Factors associated with occupational disability classification”

**2014:26** Angelov Nikolay and Marcus Eliason “The differential earnings and income effects of involuntary job loss on workers with disabilities”

**2014:27** Angelov Nikolay and Marcus Eliason “The effects of targeted labour market programs for job seekers with occupational disabilities”

**2014:28** Carlsson Mikael, Julián Messina and Oskar Nordström Skans “Firm-level shocks and labor adjustments”

**2014:29** Lindgren Karl-Oskar, Sven Oskarsson and Christopher T. Dawes “Can political inequalities be educated away? Evidence from a Swedish school reform”

**2015:1** Avdic Daniel “A matter of life and death? Hospital distance and quality of care: evidence from emergency hospital closures and myocardial infarctions”

**2015:2** Eliason Marcus “Alcohol-related morbidity and mortality following involuntary job loss”

**2015:3** Pingel Ronnie and Ingeborg Waernbaum “Correlation and efficiency of propensity score-based estimators for average causal effects”

### Dissertation series

**2014:1** Avdic Daniel “Microeconometric analyses of individual behaviour in public welfare systems”

**2014:2** Karimi Arizo “Impacts of policies, peers and parenthood on labor market outcomes”

**2014:3** Eliasson Tove “Empirical essays on wage setting and immigrant labor market opportunities”

**2014:4** Nilsson Martin “Essays on health shocks and social insurance”

**2014:5** Pingel Ronnie “Some aspects of propensity score-based estimators for causal inference”

**2014:6** Karbownik Krzysztof “Essays in education and family economics”