

Chetverikov, Denis N.; Wilhelm, Daniel

**Working Paper**

## Inference for rank-rank regressions

cemmap working paper, No. CWP11/24

**Provided in Cooperation with:**

Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Chetverikov, Denis N.; Wilhelm, Daniel (2024) : Inference for rank-rank regressions, cemmap working paper, No. CWP11/24, Centre for Microdata Methods and Practice (cemmap), London,  
<https://doi.org/10.47004/wp.cem.2024.1124>

This Version is available at:

<https://hdl.handle.net/10419/300409>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Inference for rank-rank regressions

---

Denis Chetverikov  
Daniel Wilhelm

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP11/24

# Inference for Rank-Rank Regressions

Denis Chetverikov\*

Daniel Wilhelm†

May 22, 2024

## Abstract

Slope coefficients in rank-rank regressions are popular measures of intergenerational mobility. In this paper, we first point out two important properties of the OLS estimator in such regressions: commonly used variance estimators do not consistently estimate the asymptotic variance of the OLS estimator and, when the underlying distribution is not continuous, the OLS estimator may be highly sensitive to the way in which ties are handled. Motivated by these findings we derive the asymptotic theory for the OLS estimator in a general rank-rank regression specification without making assumptions about the continuity of the underlying distribution. We then extend the asymptotic theory to other regressions involving ranks that have been used in empirical work. Finally, we apply our new inference methods to three empirical studies. We find that the confidence intervals based on estimators of the correct variance may sometimes be substantially shorter and sometimes substantially longer than those based on commonly used variance estimators. The differences in confidence intervals concern economically meaningful values of mobility and thus may lead to different conclusions when comparing mobility across different regions or countries.

---

\*Department of Economics, University of California at Los Angeles, 315 Portola Plaza, Bunche Hall, Los Angeles, CA 90024, USA; E-Mail address: [chetverikov@econ.ucla.edu](mailto:chetverikov@econ.ucla.edu).

†Departments of Statistics and Economics, Ludwig-Maximilians-Universität München, Akademiestr. 1, 80799 Munich, Germany; E-Mail address: [d.wilhelm@lmu.de](mailto:d.wilhelm@lmu.de). The author gratefully acknowledges financial support from the European Research Council (Starting Grant No. 852332). We thank Owen Groves, Philipp Kropp, Pawel Morgen, and Nicholas Vieira for excellent research assistance.

# 1 Introduction

Regressions involving ranks are widely used in empirical work in economics. A notable example is a rank-rank regression for measuring the persistence in socioeconomic status across generations. [Dahl and DeLeire \(2008\)](#) and [Chetty et al. \(2014\)](#) have been influential in promoting this approach and a vast and fast-growing empirical literature is using such and related regressions for the study of intergenerational mobility in different socioeconomic outcomes, countries, regions, and time periods.<sup>1</sup> The resulting measures are crucial inputs to broader political and public debates about inequality, about the importance of the family and the neighborhood into which children are born, and about how to create opportunities for children to rise out of poverty ([Mogstad and Torsvik \(2023\)](#)). Beyond intergenerational mobility, regressions involving ranks are used in a range of other areas such as behavioral economics (e.g., [Huffman et al. \(2022\)](#)), development (e.g., [Bagde et al. \(2016\)](#)), education (e.g., [Murphy and Weinhardt \(2020\)](#)), health (e.g., [Grönqvist et al. \(2020\)](#)), industrial organization (e.g., [Cagé et al. \(2019\)](#)), labor (e.g., [Faia and Pezone \(2023\)](#)), migration (e.g., [Ward \(2022\)](#)), and urban economics (e.g., [Lee and Lin \(2017\)](#)).

In its simplest form, a rank-rank regression in the intergenerational mobility literature consists of performing two steps: first, rank a child’s and their parent’s socioeconomic status (say, income) in their respective distributions and, second, run a regression of the child’s rank on a constant and the parent’s rank. The slope coefficient in this regression is then interpreted as a relative measure of intergenerational mobility. A small value of the slope indicates low dependence of the child’s position in their income distribution on the parent’s position in their income distribution, and thus high mobility. Applied work often assumes that the slope coefficient is equal to the rank correlation of the two incomes, which facilitates the interpretation of its magnitude as a parameter taking values between minus one and one.

In this paper, we develop an asymptotic theory for the OLS estimator of such rank-rank regressions and other regressions involving ranks.

First, we document two important properties of the OLS estimator in rank-rank regressions: commonly used standard errors are not valid and, when the underlying distribution is not continuous, then the OLS estimator may be highly sensitive to how ties in the ranks are handled. We show that the homoskedastic and Eicker-White variance estimators are not consistently estimating the asymptotic variance of the OLS estimator. In fact, their probability limits may be too large or too small relative to the true asymptotic variance, depending on the shape of the copula of the two variables to be ranked. As a consequence, inference based on these commonly used variance estimators may be conservative or fail to satisfy coverage criteria. In addition, when the underlying distribution is not continuous, then the regression estimand and the statistical properties of the OLS estimator depend, sometimes crucially, on the way in which

---

<sup>1</sup>Some examples are [Olivetti and Paserman \(2015\)](#), [Black et al. \(2019\)](#), [Abramitzky et al. \(2021\)](#), [Fagereng et al. \(2021\)](#), [Nybom and Stuhler \(2023\)](#), and [Ward \(2023\)](#). Recent surveys by [Deutscher and Mazumder \(2023\)](#) and [Mogstad and Torsvik \(2023\)](#) provide numerous further examples and references.

ties are handled.

Second, we derive the asymptotic distribution of the OLS estimator in a general rank-rank regression with covariates and without assumptions on whether the underlying distribution is continuous or not. In the special case in which the distributions are continuous and there are no covariates, the OLS estimator is equal to Spearman’s rank correlation, and our limiting distribution coincides with that derived by [Hoeffding \(1948\)](#) for Spearman’s rank correlation. In other regression specifications or when at least one of the marginal distributions is not continuous (e.g., because of zero incomes or top-coding), then the OLS estimator is not equal to Spearman’s rank correlation anymore and Hoeffding’s results do not apply. Importantly, in the case of non-continuous distributions, the estimand and the estimator’s asymptotic distribution are both sensitive to how ties are handled.

Third, we extend the asymptotic theory to other regressions involving ranks. For example, in the empirical literature studying intergenerational mobility, rank-rank regressions are sometimes run separately in different regions of a country, for instance in commuting zones within the U.S., but the ranks are computed in the national distribution rather than within the regions (e.g., as in [Chetty et al. \(2014\)](#), [Acciari et al. \(2022\)](#), [Corak \(2020\)](#)). This type of regression yields rank-rank slopes for each commuting zone with ranks defined on a common scale. We cover such clustered rank-rank regressions. Two further extensions treat the case in which the regression involves a general, non-ranked, outcome variable and a ranked regressor (e.g., as in [Murphy and Weinhardt \(2020\)](#)) and the case in which the regression involves a ranked outcome and a general, non-ranked regressor (e.g., as in [Abramitzky et al. \(2021\)](#)). For all these extensions, we derive the asymptotic distribution of the OLS estimator.

Fourth, for each of the regression models involving ranks, we propose a consistent estimator of the true asymptotic variance of the OLS estimator. Software implementations of these are provided in the R package `csranks`, which is available on CRAN. A Stata command is under development.

Finally, we discuss three empirical applications in which we compare inference methods based on our new variance estimators to those based on the commonly used variance estimators.

The first application is to [Ward \(2023\)](#), who studies intergenerational mobility in the U.S. in terms of occupational status using PSID data. In this application, the confidence intervals based on our new variance estimators are similar to those based on commonly used variance estimators, except on a small subsample of parent-child cohort pairs in which the homoskedastic variance estimator leads to conservative inference.

In the second application, we analyze the data from [Asher et al. \(2024\)](#), who studies intergenerational mobility in India in terms of education. In this data set, the marginal distributions of education are discrete, which generates ties in the ranks. Tied observations can be assigned a range of possible ranks without changing the overall ranking. We find that the rank-rank slopes are highly sensitive to which of these possible values for the rank are chosen. If the largest possible is chosen, then the rank-rank slopes are very large (close to 1) for early birth cohorts and then decline across cohorts to a value of 0.65, still a very large estimate. If the smallest

possible value of the rank of tied observations is chosen, then the estimates remain roughly constant at a low value around 0.4 across cohorts. The rank-rank slopes are therefore so sensitive to the definition of the ranks that neither their trends across cohorts can be inferred nor can one make confident comparisons with other countries like the U.S., where a comparable estimate is 0.46. As we show, for discrete variables the rank-rank slope is not equal to Spearman’s rank correlation. Interestingly, in this application, Spearman’s rank correlation is insensitive to the definition of the ranks and roughly constant across birth cohorts at a value around 0.5. Therefore, robust conclusions about relative mobility can be reached based on Spearman’s rank correlation, but not based on rank-rank slopes.

For the third application, we design a simulation experiment calibrated to data analyzed by [Chetty et al. \(2014, 2018\)](#); [Chetty and Hendren \(2018\)](#) to study intergenerational income mobility in commuting zones in the U.S.. We find that our (correct) standard errors may be substantially larger or smaller than those computed using the homoskedastic or Eicker-White variance estimators. For some of the commuting zones, the latter standard errors may be up to 60% too large or up to 20% too small. The differences in standard errors translate into economically meaningful differences in confidence intervals. For instance, we find that mobility in several commuting zones in California are significantly different from those in Italy and Canada according to our inference method. In contrast, the homoskedastic and Eicker-White variance estimators lead to confidence intervals that are conservative and include the mobility values of Italy and Canada. Similarly, there are commuting zones in California whose mobility is not significantly different from those in Norway and Canada according to our inference method, but the homoskedastic variance is too small and leads to confidence intervals that do not include Norway’s and Canada’s mobility values. Such comparisons of mobility across regions and countries are common in the literature. To avoid misled conclusions, it may therefore be important to conduct inference based on estimators of the correct variance.

The derivations of our theoretical results are based on the theory of U-statistics, as originally developed in [Hoeffding \(1948\)](#) and comprehensively reviewed in [Serfling \(2002\)](#). Alternatively, since our estimation problem is semiparametric, we could have derived the same results using empirical process theory as, e.g., in [Newey \(1994\)](#). In addition, instead of letting the controls enter the rank-rank regression in a separately additive fashion, we could consider a more flexible specification, where the slope coefficient of the rank-rank regression depends on the controls nonparametrically. For such a model, we would be able to obtain the asymptotic distribution results using the theory of influence function adjustments, as developed in [Chernozhukov et al. \(2022\)](#), or other semiparametric techniques. However, we focus on the simple, linear rank regressions as they seem to be most relevant for empirical work.

Our paper contributes to the general literature on nonparametric rank statistics. Since the literature is large, we provide here only some key references, referring an interested reader, for example, to a recent review [Chatterjee \(2023\)](#). As mentioned above, [Hoeffding \(1948\)](#) derived the asymptotic distribution for Spearman’s rank correlation in the case of continuous distributions. [Nešlehova \(2007\)](#) defined rank correlation measures for noncontinuous distributions, using a

specific way of handling ties, with the extension satisfying several intuitive axioms that any measure of concordance between random variables should satisfy. Mesfioui and Quesy (2010) extended the results of Nešlehova (2007) to cover the case of more than two variables. Genest et al. (2013) studied the problem of estimating Spearman’s rank correlation in the multivariate case with discontinuities. Ornstein and Lyhagen (2016) derived the asymptotic normality result and the asymptotic variance formula for the Spearman’s rank correlation for variables with finite support. We note here that these extensions are, although relevant, quite different from our work. In particular, we study the OLS estimator in general regression specifications involving ranks, which coincide with the Spearman’s rank correlation only in the special case of the rank-rank regression when the marginal distributions are continuous and there are no covariates. Otherwise, our estimands and estimators differ from those in the papers listed here.

Our work is also related to Klein et al. (2020), Mogstad et al. (2024) and Bazylik et al. (2024), who also proposed inference methods involving ranks, but the key difference is that they focused on inference on ranks of population parameters whereas we consider inference on features of the distribution of ranked observations.

The paper is organized as follows. We conclude the introduction with a survey on how empirical researchers in economics use regressions involving ranks. In the next section, we show that commonly used OLS variance estimators are not consistent in the context of rank-rank regressions and discuss how pointmasses in the marginal distributions affect inference. In Section 3, we derive a general asymptotic theory for rank-rank regressions. In Section 4, we provide three extensions, covering other regressions involving ranks which have been used in empirical work. In Section 5, we present three empirical applications. All proofs are contained in the online appendix.

## 1.1 Survey on the Use of Regressions Involving Ranks

Empirical researchers often transform variables into ranks before using them in a regression. To document this practice, we used Google Scholar to search for articles published between January 2013 and February 2024 containing the words “rank” and “regression”.<sup>2</sup> We restricted the search to the journals American Economic Review (excluding comments, P&P), Journal of Political Economy (excluding JPE Micro and JPE Macro), Quarterly Journal of Economics, and Review of Economic Studies. We dropped all theoretical papers without an empirical application and those that used the word “rank” in a different context, leading to a sample of 62 papers.

Many of the papers contain a large number of different regression specifications involving ranks, but we only record which types of regressions (“rank-rank”, “level-rank”, “rank-level”), which types of estimators (“OLS”, “TSLS”, “nonparametric”, “other”), and which types of standard errors occur in a paper. “Rank-rank” refers to a regression in which both the outcome and at least one regressor have been transformed into ranks before running the regression. Similarly,

---

<sup>2</sup>With this search we may have missed papers that do not use the word “rank” to describe their transformation of the variables, but rather words like “percentiles” or “deciles”.

	level-rank	rank-level	rank-rank
<i>Panel A: type of estimator</i>			
OLS	28	16	21
TSLS	3	1	1
nonparametric	24	3	12
other	7	3	0
<i>Panel B: type of standard error</i>			
homoskedastic/unknown	10	3	11
robust	7	5	0
clustered	18	6	5
other	0	2	1
none	27	7	17
total	62	23	34

**Table 1:** Number of (paper  $\times$  estimator type  $\times$  regression type  $\times$  standard error type) combinations by type of estimator and type of standard error. The sample consists of papers published between January 2013 and February 2024 in American Economic Review, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies.

“level-rank” (“rank-level”) refers to a regression in which at least one regressor (only the outcome) but not the outcome (none of the regressors) has been transformed into ranks. We obtain a total of 119 (paper  $\times$  estimator type  $\times$  regression type  $\times$  standard error type) combinations in the sample. More details about the literature search can be found in Appendix A.

Table 1 categorizes the 119 observations by type of estimator (Panel A) and by type of standard error (Panel B). First, consider the 34 observations involving rank-rank specifications. The majority (21) was estimated by OLS. Many papers (10) did not specify which method for the computation of standard errors was used. In most of these, we suspected the use of homoskedastic variance estimators and thus grouped them together with the one paper that explicitly indicated “homoskedastic” standard errors (“homoskedastic/unknown”). Many regression results were presented without standard errors (17). Not a single paper employed robust standard errors. Among the papers employing at least one rank-rank regression, the majority (about 75%) defined ranks in the sample in which the regression was run (as in Section 3 below). The remaining 25% employed clustered regressions (as in Section 4.1 below), in which ranks are defined in a larger population.

Level-rank regressions occurred in 62 of the observations, most of them employing OLS (28), but a substantial share (24) employed nonparametric methods. For these regressions, clustering or not reporting standard errors was most common. Rank-level regressions occurred somewhat less frequently in our sample (23).



To gain some insight into the topics of the papers Figure 14 in the appendix shows the number of times each top-level JEL-code appears on the papers that contain regressions involving ranks. Labor and Demographic Economics (JEL codes starting with “J”) are the most frequent, but health, education, public economics, and microeconomics are also listed as classifications by a substantial share of the papers. Overall, the graph indicates that regressions involving ranks are used in variety of subfields of economics.

## 2 Motivation

To motivate the subsequent theoretical developments, in this section we focus on rank-rank regressions and show that commonly used inference methods for these regressions are not valid, even in the simplest case when the underlying variables are continuously distributed and there are no covariates. Then, we discuss how the presence of pointmasses affects the estimand and the statistical properties of the OLS estimator.

### 2.1 Rank-Rank Regressions With Continuous Marginal Distributions

For concreteness, consider a child’s income  $Y$  and their parent’s income  $X$ . Suppose we are interested in measuring intergenerational mobility by running a regression of the child’s income rank on a constant and the parent’s income rank. The slope coefficient from this regression reflects the persistence of the two generations’ positions in their respective income distributions. A small value of the slope indicates low persistence and thus high mobility.

To simplify the discussion we start with the simplest possible rank-rank regression in which there are no covariates and the two income distributions are continuous. Let  $F$  be the joint distribution of  $(X, Y)$  and let  $F_X$  and  $F_Y$  be the corresponding marginals. Given an i.i.d. sample  $\{(X_i, Y_i)\}_{i=1}^n$  from  $F$ , we run the rank-rank regression as follows. First, compute the child’s income rank  $R_i^X := \widehat{F}_X(X_i)$  and the parent’s income rank  $R_i^Y := \widehat{F}_Y(Y_i)$ , where  $\widehat{F}_X$  and  $\widehat{F}_Y$  denote the empirical cdfs of  $X$  and  $Y$ . Then, run a regression of  $R_i^Y$  on a constant and  $R_i^X$ . The OLS estimator of the slope parameter can then be written as the sample covariance of  $R_i^Y$  and  $R_i^X$  divided by the sample variance of  $R_i^X$ :

$$\widehat{\rho} := \frac{\sum_{i=1}^n (R_i^Y - \bar{R}^Y)(R_i^X - \bar{R}^X)}{\sum_{i=1}^n (R_i^X - \bar{R}^X)^2},$$

where  $\bar{R}^X$  and  $\bar{R}^Y$  are sample averages of  $R_i^X$  and  $R_i^Y$ , respectively. Inference is then performed by computing standard errors and confidence intervals based on the usual variance formulas for the OLS estimator. Popular choices are the homoskedastic and robust variance formulas,

$$\widehat{\sigma}_{hom}^2 := \frac{\frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i^2}{\frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2} \quad \text{and} \quad \widehat{\sigma}_{EW}^2 := \frac{\frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i^2 (R_i^X - \bar{R}^X)^2}{\left(\frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2\right)^2},$$

where  $\widehat{\varepsilon}_i := R_i^Y - \widehat{\beta} - \widehat{\rho}R_i^X$  for all  $i = 1, \dots, n$  and  $\widehat{\beta} := \bar{R}^Y - \widehat{\rho}\bar{R}^X$ . The estimator  $\widehat{\sigma}_{hom}^2$  is sometimes used when regression errors are believed to be homoskedastic whereas  $\widehat{\sigma}_{EW}^2$  is robust to heteroskedasticity and is often referred to as the Eicker-White estimator.

The OLS estimator  $\hat{\rho}$  is also often interpreted as Spearman's rank correlation,

$$\hat{\rho}_S := \frac{\sum_{i=1}^n (R_i^Y - \bar{R}^Y)(R_i^X - \bar{R}^X)}{\sqrt{\sum_{i=1}^n (R_i^Y - \bar{R}^Y)^2 \sum_{i=1}^n (R_i^X - \bar{R}^X)^2}}. \quad (1)$$

Indeed, given that the marginal distributions  $F_X$  and  $F_Y$  are both assumed to be continuous, we have

$$\hat{\rho} = \hat{\rho}_S \sqrt{\frac{\sum_{i=1}^n (R_i^Y - \bar{R}^Y)^2}{\sum_{i=1}^n (R_i^X - \bar{R}^X)^2}} = \hat{\rho}_S, \quad (2)$$

with probability one.<sup>3</sup> Thus, the OLS estimator of the rank-rank regression and Spearman's rank correlation are numerically identical in finite samples. As a result, they must have the same asymptotic properties, which we now discuss.

First, it is well known that the probability limit of  $\hat{\rho}_S$  is

$$\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 = \text{Corr}(F_X(X), F_Y(Y)), \quad (3)$$

with  $C$  denoting the copula of the pair  $(X, Y)$ , i.e.  $C(x, y) := P(F_X(X) \leq x, F_Y(Y) \leq y)$  for all  $x, y \in [0, 1]$ . Second, [Hoeffding \(1948\)](#) showed that

$$\sqrt{n}(\hat{\rho}_S - \rho) \rightarrow_d N(0, \sigma^2),$$

where

$$\sigma^2 := 9 \text{Var} \left( (2F_X(X) - 1)(2F_Y(Y) - 1) + 4\psi_X(X) + 4\psi_Y(Y) \right) \quad (4)$$

and

$$\psi_X(x) := \int (F(x, y) - F_X(x)F_Y(y)) dF_Y(y), \quad x \in \mathbb{R}, \quad (5)$$

$$\psi_Y(y) := \int (F(x, y) - F_X(x)F_Y(y)) dF_X(x), \quad y \in \mathbb{R}. \quad (6)$$

The main goal of this subsection is to point out that the OLS variance estimators  $\hat{\sigma}_{hom}^2$  and  $\hat{\sigma}_{EW}^2$ , which are commonly used in applied work, are not consistently estimating the true asymptotic variance  $\sigma^2$ . In consequence, inference based on these variance estimators may be misleading.

To show that the estimators  $\hat{\sigma}_{hom}^2$  and  $\hat{\sigma}_{EW}^2$  are not consistent for  $\sigma^2$ , we first derive probability limits for both estimators. To do so, define

$$M_{kl} := E[(F_X(X) - 1/2)^k (F_Y(Y) - 1/2)^l], \quad k, l \in \{0, 1, 2, 3, 4\}.$$

The following lemma provides the probability limits.

**Lemma 1.** *Let  $\{(X_i, Y_i)\}_{i=1}^n$  be an i.i.d. sample from a distribution  $F$  with continuous marginals. Then*

$$\sigma_{hom}^2 := \text{plim}_{n \rightarrow \infty} \hat{\sigma}_{hom}^2 = 1 - \rho^2 \quad (7)$$

---

<sup>3</sup>To see this, note that by continuity of the distributions, with probability one there are no ties in the data and  $R_1^Y, \dots, R_n^Y$  is just a reordering of  $R_1^X, \dots, R_n^X$ .

and

$$\sigma_{EW}^2 := \text{plim}_{n \rightarrow \infty} \hat{\sigma}_{EW}^2 = 144 \left( M_{22} - 2\rho M_{31} + \frac{\rho^2}{80} \right). \quad (8)$$

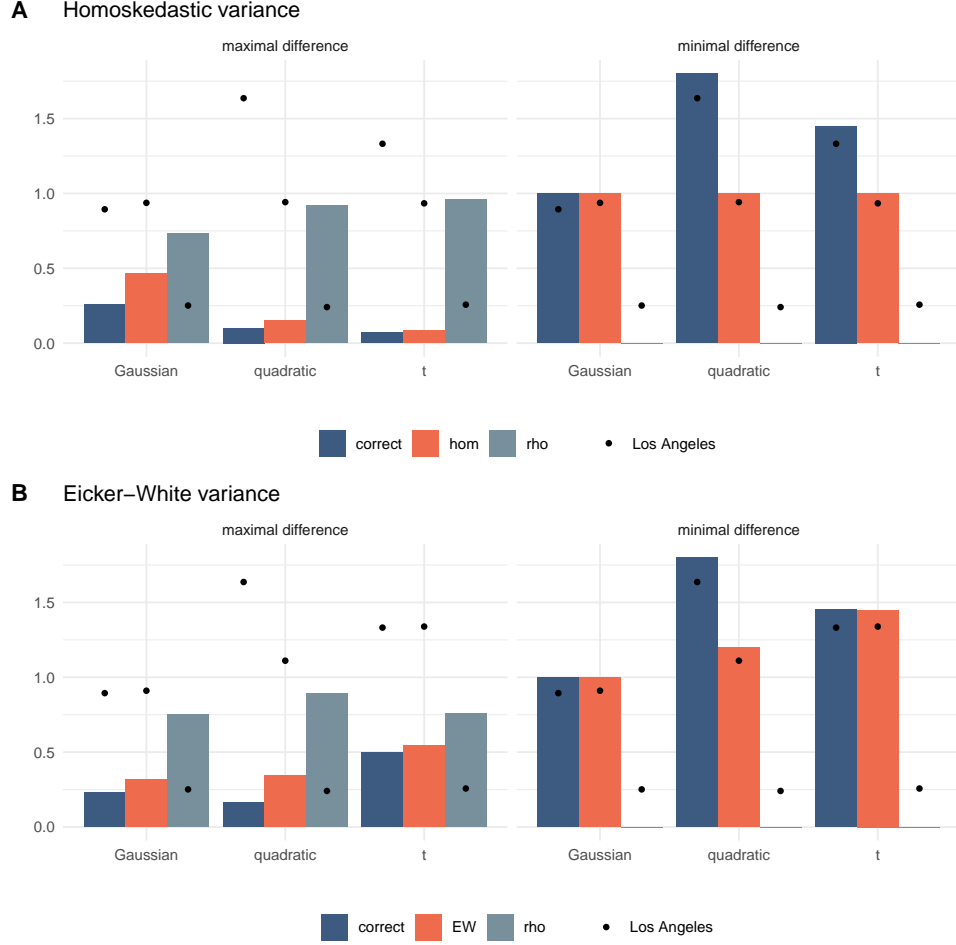
The proof of this result, as well as of all other results presented in this paper, can be found in Online Appendix E.

An important special case of this lemma occurs when  $X$  and  $Y$  are independent, in which case  $\rho = 0$  and both probability limits are equal to the correct variance,  $\sigma_{hom}^2 = \sigma_{EW}^2 = \sigma^2 = 1$ . In general, however, the three variances are different. In fact,  $\sigma_{hom}^2$  and  $\sigma_{EW}^2$  can each be larger or smaller than the correct variance  $\sigma^2$ . Appendix B.1 shows that all three variances depend on the joint distribution  $F$  only through its copula  $C$ . Therefore, whether and by how much the variances are larger or smaller depends on the shape of the copula of child and parent incomes. To illustrate this point we compare the three variances in three simple parametric families of copulas.

Specifically, we compute the minimal and maximal differences  $\sigma_{hom}^2 - \sigma^2$  and  $\sigma_{EW}^2 - \sigma^2$  over three families of copulas: the Gaussian, the Student-t (“t”), and a “quadratic” copula. In all cases, we optimize over the parameters of the copulas so as to find the largest and smallest difference in variances. Details of the specifications and solutions of the optimization problems can be found in Appendix B.2.

Panels A and B in Figure 1 show the variances achieving the minimal and maximal differences for the homoskedastic (“hom”) and the Eicker-White (“EW”) variances within the different families of copulas. They also display the rank correlation  $\rho$  of the copulas that achieve the optima. Within the t and the quadratic families, the correct variance  $\sigma^2$  may be substantially larger than the hom and EW variances, up to about 70% larger. Within the Gaussian family this is not possible, but the correct variance may be significantly smaller than the other two, almost half as large. Within the quadratic and the t families, the correct variance may also be smaller than the other two, but to a lesser extent than in the Gaussian family.

Interestingly, the minimal difference in variances is achieved for copulas that have rank correlation close to zero whereas the maximal differences occur for copulas with high rank correlation. To show that substantial differences in variances occur not only for very small or very large rank correlations, the graphs also indicate the value of the three different variances for the copula that matches the rank correlation of child and parent incomes in Los Angeles, one of the commuting zones that we further analyze in Section 5.3. In the data by Chetty et al. (2022a), Los Angeles has a rank correlation of 0.254. Since individual-level incomes are not publicly available, we do not know the true copula of incomes in Los Angeles. The dots in Figure 1 indicate the values of the correct, the hom, and the EW variances if the copula were Gaussian, t, or quadratic. If it were Gaussian, the hom and EW variances would be slightly larger than the correct one. If it were quadratic, however, the hom and EW variances would both be substantially smaller than the correct variance. Finally, in the t family, the hom variance would be substantially smaller than the correct variance, but the EW variance would be about the same as the correct one.



**Figure 1:** Variances achieving minimal and maximal differences within different families of copulas. Specifications and details can be found in Appendix B. “correct” refers to  $\sigma^2$ , “hom” to  $\sigma_{hom}^2$ , “EW” to  $\sigma_{EW}^2$ , and “rho” to the rank correlation of the copula achieving the optimized difference. The dots indicate the rank correlation in Los Angeles and the implied variances for the particular copula within the parametric family that corresponds to that rank correlation.

The variances  $\sigma^2$ ,  $\sigma_{hom}^2$ , and  $\sigma_{EW}^2$  are all bounded, and so mutual differences between them are bounded as well. This is not true for ratios, however. The following lemma shows that if we do not restrict the distributions  $F$ , the ratios  $\sigma_{hom}^2/\sigma^2$  and  $\sigma_{EW}^2/\sigma^2$  can both be arbitrarily large. To emphasize the dependence of the variances on the joint distribution  $F$  of  $(X, Y)$ , we use the notation  $\sigma_{hom}^2(F)$ ,  $\sigma_{EW}^2(F)$ , and  $\sigma^2(F)$ .

**Lemma 2.** *There exists a sequence  $\{F_k\}_{k \geq 1}$  of distributions on  $\mathbb{R}^2$  with continuous marginals such that  $\sigma_{hom}^2(F_k)/\sigma^2(F_k) \rightarrow \infty$  and  $\sigma_{EW}^2(F_k)/\sigma^2(F_k) \rightarrow \infty$  as  $k \rightarrow \infty$ .*

An implication of this lemma is that the two variances  $\sigma_{hom}^2$  and  $\sigma_{EW}^2$  can be larger than the correct variance by arbitrarily large factors. As the proof of the lemma reveals, the divergence occurs under sequences of copulas approaching perfect dependence so that both variances tend to zero, but the correct variance converges to zero at a faster rate than the variances  $\sigma_{hom}^2$  and

$\sigma_{EW}^2$ .<sup>4</sup>

Given the result in Lemma 2, the next interesting question is how small the variances  $\sigma_{hom}^2$  and  $\sigma_{EW}^2$  can be relative to the true variance  $\sigma^2$ . The following lemma provides a partial answer to this question in the case of  $\sigma_{hom}^2$ .

**Lemma 3.** *There exists a constant  $c > 0$  such that  $\sigma_{hom}^2(F)/\sigma^2(F) \geq c$  for all distributions  $F$  on  $\mathbb{R}^2$  with continuous marginals.*

By this lemma, the ratio  $\sigma_{hom}^2/\sigma^2$  is bounded away from zero, so that  $\sigma_{hom}^2$  can be smaller than  $\sigma^2$  only by a factor that is bounded from below by the constant  $c$ .<sup>5</sup> However, it is important to emphasize that this constant may be rather small. In particular, this constant is strictly smaller than one, as there do exist distributions  $F$  such that  $\sigma_{hom}^2(F)/\sigma^2(F) < 1$ . For example, we have already seen in Figure 1 that for simple well-known classes of copulas, like the t copula, the hom variance may be substantially smaller than the correct variance. In the empirical applications of Section 5, we also find that the hom variance may be substantially smaller than the correct variance.

In conclusion, this subsection has shown that commonly used formulas for the estimation of asymptotic variances, and thus also for standard errors, do not yield the correct asymptotic variance of the OLS estimator in a rank-rank regression. In particular, the homoskedastic and the Eicker-White standard errors may be too small or too large depending on the shape of the copula. In consequence, confidence intervals may be too short or too wide, possibly leading to under-coverage or conservative inference.

## 2.2 Rank-Rank Regressions With Noncontinuous Marginal Distributions

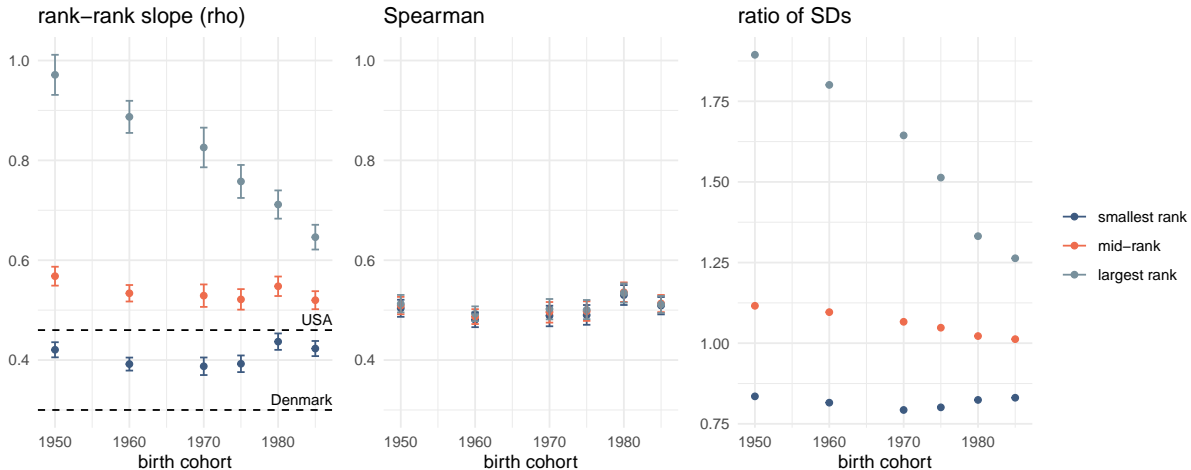
The well-established asymptotic theory for Spearman’s rank correlation described in the previous subsection crucially depends on the assumption that both marginals of the distribution  $F$  are continuous. In empirical applications of rank-rank regressions, however, pointmasses are common. For instance, incomes may be top-coded and there may be pointmasses at zero or negative incomes (e.g. as in Chetty and Hendren (2018)).<sup>6</sup> In addition, ranks may be computed from discrete measures other than income, e.g., occupational status as in Ward (2023), human capital as in de la Croix and Goni (2022), or years of education as in Asher et al. (2024). In such cases, pointmasses in the marginal distributions (of  $Y$  and  $X$ ) create ties in the ranks and their presence changes inference in the rank-rank regression in at least three important ways: (i)

---

<sup>4</sup>The copulas leading to large ratios of the variances are not the same as those leading to large differences between the variances. The large differences in Figure 1 are achieved when at least one of the variances is far from zero whereas large ratios are achieved when both are close to zero. This explains why larger ratios than those implied by Figure 1 are possible.

<sup>5</sup>We do not know whether a version of Lemma 3 holds for the EW variance  $\sigma_{EW}^2$  but we would be surprised if that were not the case.

<sup>6</sup>In fact, one of the commonly cited (Deutscher and Mazumder (2023), Mogstad and Torsvik (2023)) advantages of studying the rank-rank relationship over other approaches such as estimating intergenerational elasticities is that the rank-based approach is applicable in the presence of zero incomes.



**Figure 2:** Estimates of the rank-rank slope  $\rho$ , and Spearman’s rank correlation by birth cohort and with 95% marginal confidence intervals. The colors refer to different definition of ranks: “smallest rank” ( $\omega = 0$ ), “mid-rank” ( $\omega = 0.5$ ), and “largest rank” ( $\omega = 1$ ). The dashed-lines indicate similar measures of mobility for the U.S. and Denmark as reported in Asher et al. (2024).

the interpretation of the estimand in the rank-rank regression depends on whether there are ties or not, (ii) the estimand varies with the way in which ties are handled, and (iii) the statistical properties of the OLS estimator differ for different ways of handling ties. We now illustrate each of these three points in more detail.

**Interpretation of the estimand.** In contrast to the continuous case, when there are point-masses in the marginal distributions, the OLS estimator of the rank-rank slope is not equal to Spearman’s rank correlation. This is because, as shown in (2), the rank-rank slope is equal to Spearman’s rank correlation multiplied by the ratio of standard deviations of the ranks. For continuous distributions, this ratio is equal to one, but in the presence of pointmasses the ratio can take any value in  $(0, \infty)$ . Therefore, the slope coefficient of the rank-rank regression may lie outside of the  $[-1, 1]$  interval and cannot be interpreted as a correlation. Researchers should therefore be careful about whether they want to estimate Spearman’s rank correlation or the rank-rank slope, and potentially adjust the rank-rank slope by the ratio of standard deviations.

Figure 2 illustrates this point. It is based on the data used in Section 5.2 in which we study intergenerational mobility in India, using the data from Asher et al. (2024). In this case,  $Y$  and  $X$  are (discrete) measures of years of education of a child and their parent, respectively. The left-most panel shows estimates of the slope in a rank-rank regression for different birth cohorts of children. These are estimates, not the estimands, but since the 95% confidence intervals (based on our asymptotic theory in Section 3.1) are fairly narrow, the estimates provide insights into the possible values of the corresponding estimands. Consider the estimates labelled “largest ranks”, which refers to the constructions of ranks exactly as in Section 2.1 above. The panel in the middle shows the corresponding estimates of Spearman’s rank correlation, and the right-most panel shows the ratio of standard deviations from (2). Here, the differences between Spearman’s

rank correlation and the rank-rank slopes are striking: the former is about 0.5 for all birth cohorts while the rank-rank slope starts out at a value close to one for the 1950 birth cohort and then declines to about 0.65 for the latest birth cohort. The reason for this difference is due to the ratio of standard errors declining across birth cohorts. If Spearman’s rank correlation for the 1950 cohort was a little higher, then multiplying it by the ratio of standard deviations, which is close to 1.9, would lead to a rank-rank slope larger than one. This latter point illustrates the fact that the rank-rank slope cannot be interpreted as a correlation when there are pointmasses in the marginal distributions.

**Sensitivity of the estimand to the definition of ranks.** In the presence of pointmasses in the marginal distributions, one can assign different ranks to tied observations without changing the overall ranking. For instance, when 10 out of 100 observations are tied at the smallest value, then one can assign each of them the rank 0.01 (“smallest rank”), all of them the rank 0.1 (“largest rank”), or any value in between. The coefficients in the rank-rank regression depend on the specific way in which ranks are assigned to such tied observations.

To illustrate this point, consider the left-most panel in Figure 2 which shows the estimates of the rank-rank slopes for three different definitions of the rank. As mentioned above these are estimates, not the estimands, but since the 95% confidence intervals are fairly narrow, the estimates provide insights into the possible values of the corresponding estimands. The sensitivity of the rank-rank slope to the definition of the rank is striking. When the largest ranks are employed, the rank-rank slope is very large and declining across cohorts, but for the smallest ranks it takes small values and remains roughly constant across cohorts. Compared to a similar rank-rank slope based on a more continuous measure of education in the U.S. (dashed line), the rank-rank slopes in India may be larger or smaller depending on which definition of the ranks is used.

The differences between the rank-rank slope and Spearman’s rho are also substantial: whereas the first panel showed substantial sensitivity of the rank-rank slope to the definition of the ranks, the Spearman’s rank correlation shown in the second panel is about 0.5 for all birth cohorts and all definitions of the ranks. Equation (2) explains the reason for these difference: the rank-rank slope is equal to Spearman’s rank correlation multiplied by the ratio of standard deviations of the ranks. The third panel in Figure 2 shows the estimates of the ratios. Because the marginal distributions of education vary substantially across birth cohorts, the ratios of the standard deviations across birth cohorts. In addition, these ratios of standard deviations are very sensitive to the way in which ties in the ranks are handled. In conclusion, the rank-rank slope’s sensitivity to the definition of the ranks and its variation across birth cohorts is not inherited from Spearman’s rank correlation, but rather from the ratio of standard deviations.

**Sensitivity of the OLS estimator’s statistical properties to the definition of ranks.** In addition to the estimand, also the OLS estimator’s statistical properties depend on the way in which ties in the ranks are handled. Since the estimand is sensitive to the definition of the

ranks, so is the OLS estimator of these estimands. However, not only the probability limit and, thus, the mean of the estimator, but also its asymptotic variance depends on the definition of the ranks. This will be formally shown in Section 3.1. In consequence, standard errors for the OLS estimator vary depending on how ties in the ranks are handled. In the first panel of Figure 2 this dependence manifests itself in the length of the confidence intervals, which are substantially wider for the estimator based on the largest ranks compared to those based on the smallest ranks.

Because of the sensitivity of the OLS estimator in rank-rank regressions, we recommend that researchers check the robustness of their results with respect to the definition of the ranks.

## 2.3 Summary

The previous arguments can be summarized as follows:

1. Hoeffding (1948)'s derivation of the asymptotic distribution of Spearman's rank correlation yields the correct asymptotic distribution of the OLS estimator of the slope in a rank-rank regression in the special case with continuous marginal distributions of  $Y$  and  $X$  and without covariates.
2. Commonly used variance estimators for OLS regressions do not estimate this asymptotic variance. They may be too small or too large depending on the shape of the copula of the pair  $(X, Y)$ .
3. When at least one of the two marginal distributions of  $Y$  and  $X$  is not continuous or when there are covariates in the regression, then Hoeffding's theory does not apply. Importantly, in the noncontinuous case, the rank-rank regression estimand, the OLS estimator, and its asymptotic distribution are all sensitive to the specific way in which ties are handled.

The next section develops a general, unifying asymptotic theory for coefficients in a rank-rank regression that allows for any definition of the rank, for continuous or noncontinuous distributions, and for the presence of covariates.

# 3 Inference for Rank-Rank Regressions

## 3.1 Asymptotic Normality Result

In this section, we develop the asymptotic theory for the OLS estimator of a rank-rank regression. The results apply to continuous and non-continuous distributions  $F$ . To deal with potential ties, we consider a more general definition of the rank, in fact a class of ranks, indexed by a parameter  $\omega \in [0, 1]$ . Different definitions of the rank differ in the way in which they handle ties, but they are all equivalent in the absence of ties.



$i$	1	2	3	4	5	6	7	8	9	10
$X_i$	3	4	7	7	10	11	15	15	15	15
smallest rank: $R_i^X$ for $\omega = 0$	0.1	0.2	<b>0.3</b>	<b>0.3</b>	0.5	0.6	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>
mid-rank: $R_i^X$ for $\omega = 0.5$	0.1	0.2	<b>0.35</b>	<b>0.35</b>	0.5	0.6	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>
largest rank: $R_i^X$ for $\omega = 1$	0.1	0.2	<b>0.4</b>	<b>0.4</b>	0.5	0.6	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

**Table 2:** example of three different definitions of ranks for a given sample. Bold columns indicate ties.

For a fixed, user-specified  $\omega \in [0, 1]$ , let  $F_X^-(x) := P(X < x)$ ,  $I(u, v) := \omega \mathbb{1}\{u \leq v\} + (1 - \omega) \mathbb{1}\{u < v\}$ , and

$$R_X(x) := \omega F_X(x) + (1 - \omega) F_X^-(x), \quad x \in \mathbb{R}.$$

For an i.i.d. sample  $X_1, \dots, X_n$  from  $F_X$ , we then define the rank of  $i$  as  $R_i^X := \widehat{R}_X(X_i)$ , where

$$\widehat{R}_X(x) := \omega \widehat{F}_X(x) + (1 - \omega) \widehat{F}_X^-(x) + \frac{1 - \omega}{n}, \quad x \in \mathbb{R}, \quad (9)$$

is an estimator of  $R_X(x)$  and  $\widehat{F}_X^-(x) = n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i < x\}$  is an estimator of  $F_X^-(x)$ . This definition of the rank is such that an individual  $i$  with a large value of  $X_i$  is assigned a large rank. If  $F_X$  is continuous, then the probability of a tie among  $X_1, \dots, X_n$  is zero and the rank  $R_i^X$  is the same for all values of  $\omega$ . If  $F_X$  is not continuous, then different choices of  $\omega$  lead to definitions of the rank that handle ties differently. For instance, if  $\omega = 1$  then  $R_i^X = \widehat{F}_X(X_i)$  and tied individuals are assigned the largest possible rank. If  $\omega = 1/2$ , then  $R_i^X$  is the mid-rank as defined in [Hoeffding \(1948\)](#), which assigns to tied individuals the average of the smallest and largest possible ranks. Finally,  $\omega = 0$  leads to a definition of the rank that assigns to tied individuals the smallest possible rank. Table 2 illustrates the different definitions of ranks in an example with ties. The quantities  $F_Y$ ,  $F_Y^-$ ,  $R_Y$  and their estimators are defined analogously for  $Y$ . We assume that  $\omega$  is the same value in  $R_X$  and  $R_Y$ , so that ranks are defined consistently for both variables  $X$  and  $Y$ . Throughout the paper, this value of  $\omega$  is fixed and chosen by the researcher, so it does not appear as argument or index anywhere.

We consider the following regression model:

$$R_Y(Y) = \rho R_X(X) + W' \beta + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \right] = 0, \quad (10)$$

where  $W$  is a  $d$ -dimensional vector of covariates,  $\varepsilon$  is noise, and  $\rho$ ,  $\beta$  are regression coefficients to be estimated. In applications,  $W$  typically includes a constant, but our theoretical results below do not require that. As already indicated in Section 2.2, the interpretation of  $\rho$  depends on whether the marginal distributions of  $Y$  and  $X$  are continuous and, if at least one of them is not, how ties are handled. However, in the model (10) there is one additional reason complicating the interpretation of  $\rho$ , namely the presence of covariates: even if  $Y$  and  $X$  have continuous marginal distributions, in the presence of covariates  $W$  that are correlated with  $X$ ,  $\rho$  cannot be interpreted as a rank correlation and, in particular, does not have to lie in the interval  $[-1, 1]$ .<sup>7</sup>

<sup>7</sup>This can be seen by writing  $\rho$  as the covariance of residuals from partialling out  $W$  from the two ranks divided

Letting  $\{(X_i, W_i, Y_i)\}_{i=1}^n$  be an i.i.d. sample from the distribution of the triplet  $(X, W, Y)$ , we study the properties of the following OLS estimator of the vector of parameters  $(\rho, \beta)'$ :

$$\begin{pmatrix} \widehat{\rho} \\ \widehat{\beta} \end{pmatrix} = \left( \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} R_i^Y. \quad (11)$$

To derive the asymptotic normality of  $\widehat{\rho}$ , we introduce the projection of  $R_X(X)$  onto the covariates:

$$R_X(X) = W'\gamma + \nu, \quad E[\nu W] = 0, \quad (12)$$

where  $\nu$  is a random variable representing the projection residual, and  $\gamma$  is a vector of parameters. Consider the following regularity conditions:

**Assumption 1.**  $\{(X_i, W_i, Y_i)\}_{i=1}^n$  is an i.i.d. sample from the distribution of  $(X, W, Y)$ .

**Assumption 2.** The vector  $W$  satisfies  $E[\|W\|^4] < \infty$  and the matrix  $E[WW']$  is non-singular.

**Assumption 3.** The random variable  $\nu$  is such that  $E[\nu^2] > 0$ .

These are standard regularity conditions underlying typical regression analysis. Assumption 3 requires that the rank  $R_X(X)$  can not be represented as a linear combination of covariates. Importantly, our regularity conditions do not require  $X$  and  $Y$  to be continuously distributed.

Under these conditions, we have the following asymptotic normality result.

**Theorem 1.** Suppose that (10)–(12) hold and that Assumptions 1–3 are satisfied. Then

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{1}{\sigma_\nu^2 \sqrt{n}} \sum_{i=1}^n \left\{ h_1(X_i, W_i, Y_i) + h_2(X_i, Y_i) + h_3(X_i) \right\} + o_P(1) \rightarrow_D N(0, \sigma^2),$$

where

$$\sigma^2 := \frac{1}{\sigma_\nu^4} E \left[ (h_1(X, W, Y) + h_2(X, Y) + h_3(X))^2 \right] \quad (13)$$

with  $\sigma_\nu^2 := E[\nu^2]$  and

$$\begin{aligned} h_1(x, w, y) &:= (R_Y(y) - \rho R_X(x) - w'\beta)(R_X(x) - w'\gamma), \\ h_2(x, y) &:= E[(I(y, Y) - \rho I(x, X) - W'\beta)(R_X(X) - W'\gamma)], \\ h_3(x) &:= E[(R_Y(Y) - \rho R_X(X) - W'\beta)(I(x, X) - W'\gamma)] \end{aligned}$$

for all  $x \in \mathbb{R}$ ,  $w \in \mathbb{R}^p$ , and  $y \in \mathbb{R}$ .

As already mentioned above, the proof of this result, as well as of all other results presented in this paper, can be found in Online Appendix E. In fact, Theorem 1 follows from a more 

---

by the variance of one of the two residuals. Then, this ratio is not equal to the correlation of the ranks after partialling out  $W$  because the standard deviations of the two residuals are not necessarily equal.

general result, which we state in Online Appendix D (Theorem 5). The more general result shows joint asymptotic normality for  $\hat{\rho}$  and  $\hat{\beta}$ ,

$$\sqrt{n} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma), \quad \Sigma := E[\psi_i \psi_i'],$$

and provides an explicit formula for  $\psi_i$ . The joint asymptotic distribution for  $\hat{\rho}$  and  $\hat{\beta}$  is useful for empirical work in the intergenerational mobility literature. Suppose there are no covariates so  $W$  only contains a constant. Besides the rank-rank slope, another popular measure of intergenerational mobility (Deutscher and Mazumder (2023)) is the expected income rank of a child given that their parent's income rank is equal to some given value  $p$ :

$$\theta_p := \beta + \rho p.$$

With the asymptotic joint distribution above, it is straightforward to compute standard errors and confidence intervals for  $\theta_p$ .

Theorem 1 has the following interpretation. If we knew the population ranks  $R_Y(Y_i)$  and  $R_X(X_i)$ , we would simply estimate  $\rho$  and  $\beta$  by an OLS regression of  $R_Y(Y_i)$  on  $R_X(X_i)$  and  $W$ . The asymptotic variance of such an estimator for  $\rho$  would be given by  $E[h_1(X, W, Y)^2]/\sigma_\nu^2$ , which is the familiar OLS variance formula. Without knowing the population ranks, however, we have to plugin their estimators, and the extra terms in the asymptotic variance, represented by  $h_2(X, Y)$  and  $h_3(X)$ , provide the adjustments necessary to take into account the noise coming from estimated ranks. Importantly, the estimation error in the ranks is of the same order of magnitude as that in the OLS estimator with known ranks. Therefore, the estimation error in the ranks does not become negligible in large samples, not even in the limit as the sample size grows to infinity.

The derivation of the asymptotic normality result uses standard techniques from the U-statistic theory. To see this consider the case without covariates so that  $W$  only contains a constant. Then, the OLS estimator of the slope can be written as

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n (R_i^Y - \bar{R}^Y)(R_i^X - \bar{R}^X)}{\frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2}.$$

It is easy to show that the denominator converges in probability to  $\sigma_\nu^2$ . The numerator is a U-statistic because it is an average of terms involving  $R_i^Y$  and  $R_i^X$ , which themselves are sample averages. Under the above assumptions one can then verify the standard conditions for such a U-statistic to be asymptotically normal (e.g., Serfling (2002)).

**Remark 1** (Comparing the variance formulas in (4) and (13)). When both  $X$  and  $Y$  are continuous random variables and  $W$  contains only a constant, the asymptotic variance formula in (13) reduces to the classical Hoeffding variance formula in (4). Indeed, replacing  $X$  and  $Y$  by  $F_X(X)$  and  $F_Y(Y)$  respectively, we can assume without loss of generality that both  $X$  and  $Y$  are  $U([0, 1])$  random variables, in which case  $\psi_X(x)$  in (5) reduces to

$$\int_0^1 (E[\mathbf{1}\{X \leq x\}\mathbf{1}\{Y \leq y\}] - xy) dy = E[\mathbf{1}\{X \leq x\}(1 - Y)] - x/2 = x/2 - E[\mathbf{1}\{X \leq x\}Y].$$

Similarly,  $\psi_Y(y)$  in (6) reduces to  $y/2 - E[\mathbb{1}\{Y \leq y\}X]$ . Hence, the variance in (4) simplifies to  $\sigma^2 = 144\text{Var}(h(X, Y))$ , where

$$h(x, y) := xy - E[\mathbb{1}\{X \leq x\}Y] - E[\mathbb{1}\{Y \leq y\}X], \quad x \in \mathbb{R}, y \in \mathbb{R}, \quad (14)$$

and it is straightforward to verify that  $h(x, y)$  coincides with  $h_1(x, w, y) + h_2(x, y) + h_3(x)$  up to an additive constant whenever  $X, Y \sim U([0, 1])$  and  $W$  contains only a constant.  $\square$

**Remark 2** (Non-continuous distributions). The asymptotic normality result in Theorem 1 holds for both continuous and noncontinuous distributions of  $Y$  and  $X$ . In the presence of pointmasses in at least one of the distributions of  $Y$  and  $X$ , the estimands  $\rho$  and  $\beta$  depend on  $\omega$ , i.e. on the way in which ties are handled. Similarly, the asymptotic variance  $\sigma^2$  of the OLS estimator derived in Theorem 1 depends on  $\omega$ . Therefore, different ways of handling ties through different choices of  $\omega$  may affect not only the estimand and the estimator, but also the value of the asymptotic variance.  $\square$

### 3.2 Consistent Estimation of the Asymptotic Variance

In this subsection, we propose an estimator of the asymptotic variance  $\sigma^2$  appearing in the asymptotic normality result in Theorem 1 and show that it is consistent. In particular, we consider the following plug-in estimator:

$$\hat{\sigma}^2 := \frac{1}{n\hat{\sigma}_\nu^4} \sum_{i=1}^n (H_{1i} + H_{2i} + H_{3i})^2,$$

where  $\hat{\sigma}_\nu^2 := n^{-1} \sum_{i=1}^n \hat{\nu}_i^2$  is an empirical analog of  $\sigma_\nu^2 = E[\nu^2]$ ,  $\hat{\nu}_i := R_i^X - W_i' \hat{\gamma}$ , and

$$\begin{aligned} H_{1i} &:= \left( R_i^Y - \hat{\rho} R_i^X - W_i' \hat{\beta} \right) \left( R_i^X - W_i' \hat{\gamma} \right), \\ H_{2i} &:= \frac{1}{n} \sum_{j=1}^n \left( I(Y_i, Y_j) - \hat{\rho} I(X_i, X_j) - W_j' \hat{\beta} \right) \left( R_j^X - W_j' \hat{\gamma} \right), \\ H_{3i} &:= \frac{1}{n} \sum_{j=1}^n \left( R_j^Y - \hat{\rho} R_j^X - W_j' \hat{\beta} \right) \left( I(X_i, X_j) - W_j' \hat{\gamma} \right), \end{aligned}$$

for all  $i = 1, \dots, n$ . The following lemma shows that this simple plug-in estimator is consistent without any additional assumptions.

**Lemma 4.** *Suppose that (10)–(12) hold and that Assumptions 1–3 are satisfied. Then  $\hat{\sigma}^2 \rightarrow_P \sigma^2$ .*

Theorem 1 and Lemma 4 give the correct way to perform inference in rank-rank regressions. For example, a  $(1 - \alpha) \times 100\%$  asymptotic confidence interval for  $\rho$  can be constructed using the standard formula

$$\left( \hat{\rho} - \frac{z_{\alpha/2} \hat{\sigma}}{\sqrt{n}}, \hat{\rho} + \frac{z_{\alpha/2} \hat{\sigma}}{\sqrt{n}} \right),$$

where  $z_{\alpha/2}$  is the number such that  $P(N(0, 1) > z_{\alpha/2}) = \alpha/2$ . Standard hypothesis testing can be performed analogously.

**Remark 3** (Bootstrapping the distribution of  $\sqrt{n}(\hat{\rho} - \rho)$ ). In addition to performing inference on  $\rho$  via a consistent estimator of the asymptotic variance  $\hat{\sigma}^2$  in Lemma 4, one can also perform inference by bootstrapping the distribution of  $\sqrt{n}(\hat{\rho} - \rho)$ . Indeed, to show bootstrap validity, observe that  $\rho = \rho(F)$  is a smooth functional (Hadamard differentiable), and our estimator  $\hat{\rho}$  takes the plug-in form,  $\hat{\rho} = \rho(\hat{F})$ . Therefore, given that the empirical distribution function  $\hat{F}$  satisfies Donsker’s theorem, it follows from Theorem 3.1 in Fang and Santos (2019) that the distribution of  $\sqrt{n}(\hat{\rho}^* - \hat{\rho})$ , where  $\hat{\rho}^*$  is the estimator  $\hat{\rho}$  obtained on a bootstrap sample, consistently estimates that of  $\sqrt{n}(\hat{\rho} - \rho)$ ; see also Bickel and Freedman (1981) for the original result on bootstrapping von Mises functionals.  $\square$

## 4 Inference for Other Regressions Involving Ranks

Motivated by regressions that have been used in empirical work, especially in the intergenerational mobility literature, we provide three extensions of the asymptotic normality result in Theorem 1. In particular, we consider (i) a rank-rank regression with clusters, where ranks are computed in the national distributions rather than in the cluster-specific distributions, (ii) a regression of a general outcome  $Y$  on the rank  $R_X(X)$ , and (iii) a regression of the rank  $R_Y(Y)$  on a covariate  $X$ . For brevity of the paper, we keep the discussions of each extension relatively short.

### 4.1 Rank-Rank Regressions With Clusters

In this subsection, we consider a population (e.g., the U.S.) that is divided into  $n_G$  subpopulations or “clusters” (e.g., commuting zones). We are interested in running rank-rank regressions separately within each cluster. The ranks, however, are computed from the distribution of the entire population (e.g., the U.S.). Chetty et al. (2014) has been influential in promoting these kind of regressions for the analysis of mobility across regions, where the scale of the mobility measure is fixed by the national distribution. The survey by Deutscher and Mazumder (2023) provides more examples of empirical work running such regressions, for instance Corak (2020) and Acciari et al. (2022).

Specifically, we consider the model

$$R_Y(Y) = \sum_{g=1}^{n_G} \mathbb{1}\{G = g\} (\rho_g R_X(X) + W' \beta_g) + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \middle| G \right] = 0 \text{ a.s.}, \quad (15)$$

where  $G$  is an observed random variable taking values in  $\{1, \dots, n_G\}$  to indicate the cluster to which an individual belongs. The quadruple  $(G, X, W, Y)$  has distribution  $F$ , and we continue to denote marginal distributions of  $X$  and  $Y$  by  $F_X$  and  $F_Y$ . The quantities  $F_X^-$ ,  $F_Y^-$ ,  $R_X(x)$ , and  $R_Y(y)$  are also as previously defined, so that  $R_X(X)$ , for instance, is the rank of  $X$  in the entire population, not the rank within a cluster. So, in model (15), the coefficients  $\rho_g$  and  $\beta_g$  are cluster-specific, but the ranks  $R_Y(Y)$  and  $R_X(X)$  are not. In consequence,  $\rho_g$  cannot

be interpreted as the rank correlation within the cluster  $g$ . Instead, in the intergenerational mobility literature, the rank-rank slope is interpreted as a relative measure of mobility in a region  $g$ , where its scale is fixed by the national population. Unlike the rank-rank slope in the model without clusters, (10), the slopes  $\rho_g$  do not only depend on the copula of  $Y$  and  $X$  in a cluster, but also on the marginal distributions of  $Y$  and  $X$  in the cluster. To see this note that adding a fixed amount to every child and parent income in a cluster does not change the ranking of children and parents within the cluster, but it may change the ranking of these individuals in their national income distributions. In conclusion, the  $\rho_g$  may then also change.

We now introduce a first-stage projection equation similar to the one in (12), except that the coefficients are cluster-specific:

$$R_X(X) = \sum_{g=1}^{n_G} \mathbb{1}\{G = g\} W' \gamma_g + \nu, \quad E[\nu W | G] = 0 \text{ a.s.} \quad (16)$$

Let  $\{(G_i, X_i, W_i, Y_i)\}_{i=1}^n$  be an i.i.d. sample from the distribution of  $(G, X, W, Y)$ . Ranks are computed using all observations, i.e.  $R_i^X := \widehat{R}_X(X_i)$  with  $\widehat{R}_X(x)$  as in (9) and  $\widehat{F}_X$  ( $\widehat{F}_X^-$ ) the (left-limit of the) empirical cdf of  $X_1, \dots, X_n$ . The computation of the rank  $R_i^Y$  is analogous.

First, notice that an OLS regression of  $R_i^Y$  on all regressors, i.e.  $(\mathbb{1}\{G_i = g\} R_i^X)_{g=1}^{n_G}$  and  $(\mathbb{1}\{G_i = g\} W_i^X)_{g=1}^{n_G}$ , produces estimates  $(\widehat{\rho}_g, \widehat{\beta}_g)_{g=1}^{n_G}$  of  $(\rho_g, \beta_g)_{g=1}^{n_G}$  that can be written as:

$$\begin{pmatrix} \widehat{\rho}_g \\ \widehat{\beta}_g \end{pmatrix} = \left( \sum_{i=1}^n \mathbb{1}\{G_i = g\} \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \mathbb{1}\{G_i = g\} \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} R_i^Y. \quad (17)$$

Therefore,  $(\widehat{\rho}_g, \widehat{\beta}_g)$  can be computed by an OLS regression of  $R_i^Y$  on  $R_i^X$  and  $W_i$  using only observations  $i$  from cluster  $g$ . Similarly,  $\widehat{\gamma}_g$  can be computed by an OLS regression of  $R_i^X$  on  $W_i$  using only observations  $i$  from cluster  $g$ . Note however, as explained above, that the ranks  $R_i^X$  and  $R_i^Y$  are computed using observations of  $X$  and  $Y$  from all clusters and thus the OLS estimators across clusters are not independent.

The following are Assumptions 1–3 adapted to the model with clusters:

**Assumption 4.**  $\{(G_i, X_i, W_i, Y_i)\}_{i=1}^n$  is a random sample from the distribution of  $(G, X, W, Y)$ .

**Assumption 5.** The vector  $W$  is such that  $E[\|W\|^4] < \infty$  and, for all  $g = 1, \dots, n_G$ , the matrix  $E[\mathbb{1}\{G = g\} W W']$  is non-singular.

**Assumption 6.** The random variable  $\nu$  is such that  $E[\mathbb{1}\{G = g\} \nu^2] > 0$  for all  $g = 1, \dots, n_G$ .

As in the previous section, note that our assumptions do not require the marginal distributions  $F_X$  and  $F_Y$  to be continuous. In addition, we introduce an assumption about the number and size of the clusters:

**Assumption 7.** The number of clusters  $n_G$  is a finite constant and  $P(G = g) > 0$  for all  $g = 1, \dots, G$ .

Observe that if the number of clusters were to increase together with the sample size  $n$ , with the number of units within each cluster being of the same order, the extra noise coming from estimated ranks would be asymptotically negligible, and the standard OLS variance formula would be applicable. The new result below applies to the case with a fixed number of clusters so that the estimation error in the ranks is not negligible even in large samples. This scenario seems reasonable, for instance, in our empirical application in which a cluster is a commuting zone within a U.S. state because each of these states contains only a small number of commuting zones relative to the sample size.

Under these four assumptions, we have the following extension of Theorem 1.

**Theorem 2.** *Suppose that (15)–(17) hold and that Assumptions 4–7 are satisfied. Then for all  $g = 1, \dots, G$ , we have*

$$\sqrt{n}(\widehat{\rho}_g - \rho_g) = \frac{1}{\sigma_{\nu,g}^2 \sqrt{n}} \sum_{i=1}^n \left\{ h_{1,g}(G_i, X_i, W_i, Y_i) + h_{2,g}(X_i, Y_i) + h_{3,g}(X_i) \right\} + o_P(1) \rightarrow_D N(0, \sigma_g^2),$$

where

$$\sigma_g^2 := \frac{1}{\sigma_{\nu,g}^4} E \left[ (h_{1,g}(G, X, W, Y) + h_{2,g}(X, Y) + h_{3,g}(X))^2 \right]$$

with  $\sigma_{\nu,g}^2 := E[\mathbb{1}\{G = g\}\nu^2]$  and

$$\begin{aligned} h_{1,g}(\bar{g}, x, w, y) &:= \mathbb{1}\{\bar{g} = g\} (R_Y(y) - \rho_g R_X(x) - w' \beta_g) (R_X(x) - w' \gamma_g), \\ h_{2,g}(x, y) &:= E[\mathbb{1}\{G = g\} (I(y, Y) - \rho_g I(x, X) - W' \beta_g) (R_X(X) - W' \gamma_g)], \\ h_{3,g}(x) &:= E[\mathbb{1}\{G = g\} (R_Y(Y) - \rho_g R_X(X) - W' \beta_g) (I(x, X) - W' \gamma_g)] \end{aligned}$$

for all  $\bar{g} \in \{1, \dots, n_G\}$ ,  $x \in \mathbb{R}$ ,  $w \in \mathbb{R}^p$ , and  $y \in \mathbb{R}$ .

In Online Appendix D, we provide a joint asymptotic normality result for all regression coefficients. Letting  $\widehat{\rho} := (\widehat{\rho}_1, \dots, \widehat{\rho}_{n_G})'$ ,  $\rho := (\rho_1, \dots, \rho_{n_G})'$ ,  $\widehat{\beta} := (\widehat{\beta}'_1, \dots, \widehat{\beta}'_{n_G})'$ , and  $\beta := (\beta'_1, \dots, \beta'_{n_G})'$ , the appendix shows that

$$\sqrt{n} \begin{pmatrix} \widehat{\rho} - \rho \\ \widehat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma), \quad \Sigma := E[\psi_i \psi_i'], \quad (18)$$

and provides an explicit formula for  $\psi_i$ . From this result, one can then easily calculate the asymptotic distribution of linear combinations of parameters. For instance, similarly as in the rank-rank regression without clusters, a popular measure of intergenerational mobility (Deutscher and Mazumder (2023)) is the expected rank of a child with parents at a given income rank  $p$ ,

$$\theta_{g,p} := \beta_g + \rho_g p. \quad (19)$$

The asymptotic distribution in (18) allows us to construct a confidence interval for  $\theta_{g,p}$  for a specific commuting zone  $g$  or simultaneous confidence sets across all commuting zones.

An estimator of the asymptotic variance  $\sigma^2$  is

$$\hat{\sigma}_g^2 := \frac{1}{n\hat{\sigma}_{\nu,g}^4} \sum_{i=1}^n (H_{1,g,i} + H_{2,g,i} + H_{3,g,i})^2,$$

where

$$\begin{aligned} \hat{\sigma}_{\nu,g}^2 &:= \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{G_i = g\} (R_i^X - W_i' \hat{\gamma}_g)^2, \\ H_{1,g,i} &:= \mathbb{1}\{G_i = g\} \left( R_i^Y - \hat{\rho}_g R_i^X - W_i' \hat{\beta}_g \right) (R_i^X - W_i' \hat{\gamma}_g), \\ H_{2,g,i} &:= \frac{1}{n} \sum_{j=1}^n \mathbb{1}\{G_j = g\} \left( I(Y_i, Y_j) - \hat{\rho}_g I(X_i, X_j) - W_j' \hat{\beta}_g \right) (R_j^X - W_j' \hat{\gamma}_g), \\ H_{3,g,i} &:= \frac{1}{n} \sum_{j=1}^n \mathbb{1}\{G_j = g\} \left( R_j^Y - \hat{\rho}_g R_j^X - W_j' \hat{\beta}_g \right) (I(X_i, X_j) - W_j' \hat{\gamma}_g) \end{aligned}$$

for all  $g \in \{1, \dots, n_G\}$  and  $i = 1, \dots, n$ . One can prove consistency of this estimator using the same arguments as those in the proof of Lemma 4.

## 4.2 Level-Rank Regressions

In this subsection, we consider a regression model with the level of  $Y$  as the dependent variable and a rank as the regressor:

$$Y = \rho R_X(X) + W' \beta + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \right] = 0. \quad (20)$$

Such a regression has been used, for instance, by [Chetty et al. \(2014\)](#) or [Abramitzky et al. \(2021\)](#) who regress a child's outcome like college attendance or teenage pregnancy on their parent's income rank. Another example is [Chetty et al. \(2011\)](#) in which wages are regressed on the rank of a test score.

For the above regression, the OLS estimator takes the form

$$\begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \left( \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} Y_i. \quad (21)$$

The following theorem derives asymptotic normality for  $\hat{\rho}$ .

**Theorem 3.** *Suppose that (12), (20), and (21) hold, that  $E[\varepsilon^4] < \infty$ , and that Assumptions 1–3 are satisfied. Then*

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{1}{\sigma_{\nu}^2 \sqrt{n}} \sum_{i=1}^n \left\{ h_1(X_i, W_i, Y_i) + h_2(X_i) + h_3(X_i) \right\} + o_P(1) \rightarrow_D N(0, \sigma^2),$$

where

$$\sigma^2 := \frac{1}{\sigma_{\nu}^4} E \left[ (h_1(X, W, Y) + h_2(X) + h_3(X))^2 \right]$$



with  $\sigma_\nu^2 := E[\nu^2]$  and

$$\begin{aligned} h_1(x, w, y) &:= (y - \rho R_X(x) - w'\beta)(R_X(x) - w'\gamma), \\ h_2(x) &:= E[Y - \rho I(x, X) - W'\beta](R_X(X) - W'\gamma), \\ h_3(x) &:= E[(Y - \rho R_X(X) - W'\beta)(I(x, X) - W'\gamma)] \end{aligned}$$

for all  $x \in \mathbb{R}$ ,  $w \in \mathbb{R}^p$ , and  $y \in \mathbb{R}$ .

In Online Appendix D, we provide a joint asymptotic normality result for all regression coefficients, i.e. we show that

$$\sqrt{n} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma), \quad \Sigma := E[\psi_i \psi_i'],$$

and derive the expression for  $\psi_i$ . A consistent estimator of the asymptotic variance can be obtained by the plugin method, analogously to our discussion in Section 3.

### 4.3 Rank-Level Regressions

In this subsection, we consider a regression model in which the outcome variable has been transformed into a rank, but the regressors are included as levels:

$$R_Y(Y) = W'\beta + \varepsilon, \quad E[\varepsilon W] = 0, \quad (22)$$

where, for simplicity of notation, we let the vector  $W$  absorb the regressor  $X$ . Examples of such a regressions in applied work are Sharkey and Torrats-Espinosa (2017) who regress the income rank on the crime rate and Ghosh et al. (2023) who regress income rank on a binary indicator for whether a couple is in an isonymous marriage.

Denote by  $W_l$  the  $l$ -th element of  $W := (W_1, \dots, W_p)'$  and by  $W_{-l}$  the vector of all elements of  $W$  except the  $l$ -th. The projection of  $W_l$  onto  $W_{-l}$  for any  $l = 1, \dots, p$  now takes the following form:

$$W_l = W_{-l}'\gamma_l + \nu_l, \quad E[\nu_l W_{-l}] = 0, \quad (23)$$

where  $\gamma_1, \dots, \gamma_p$  are  $(p-1)$ -dimensional vectors of parameters. For this model, the OLS estimator takes the following form:

$$\hat{\beta} = \left( \sum_{i=1}^n W_i W_i' \right)^{-1} \sum_{i=1}^n W_i R_i^Y. \quad (24)$$

We then have the following result.

**Theorem 4.** *Suppose that (22)–(24) hold and that Assumptions 1 and 2 are satisfied. Then*

$$\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma),$$

where  $\Sigma := E[\psi_i \psi_i']$ ,  $\psi_i := (\phi_1(W_i, Y_i), \dots, \phi_p(W_i, Y_i)')$  for all  $i = 1, \dots, n$ ,

$$\phi_l(w, y) := \frac{1}{\sigma_{\nu_l}^2} [h_{l,1}(w, y) + h_{l,2}(y)],$$

$\sigma_{\nu_l}^2 := E[\nu_l^2]$ , and

$$\begin{aligned} h_{l,1}(w, y) &:= (R_Y(y) - w' \beta)(w_l - w'_{-l} \gamma_l), \\ h_{l,2}(y) &:= E[(I(y, Y) - W' \beta)(W_l - W'_{-l} \gamma_l)], \end{aligned}$$

for all  $w \in \mathbb{R}^p$  and  $y \in \mathbb{R}$ .

The theorem shows, in particular, that for the  $l$ -th element of  $\widehat{\beta}$ , we have

$$\sqrt{n}(\widehat{\beta}_l - \beta_l) \rightarrow_D N(0, \sigma_{\beta_l}^2),$$

where

$$\sigma_{\beta_l}^2 := \frac{1}{\sigma_{\nu_l}^4} E \left[ (h_{l,1}(W, Y) + h_{l,2}(Y))^2 \right],$$

and, like above, a consistent estimator of the asymptotic variance can be obtained by the plugin method, analogously to our discussion in Section 3.

## 5 Empirical Applications

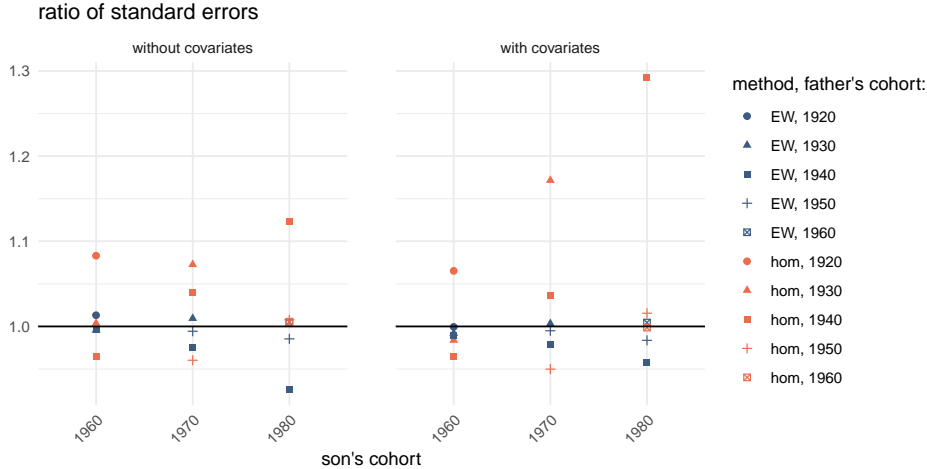
We illustrate our new inference methods in three empirical applications. Each of them analyzes intergenerational mobility, but the applications vary in the socioeconomic outcome under study and in the populations from which data are drawn. The first analyzes mobility in occupational status in the U.S., the second mobility in education in India, and the third mobility in income across U.S. neighborhoods.

### 5.1 Intergenerational Mobility in the U.S. – Occupational Status

In this section, we analyze intergenerational mobility in the U.S. as in [Song et al. \(2020\)](#) and [Ward \(2023\)](#), who measure mobility by the rank-rank relationship between fathers' and sons' occupational status. We focus on cohorts covered by the PSID dataset used in [Ward \(2023\)](#).

**Data.** We use the PSID dataset from [Ward \(2023\)](#), including his construction of the Song score (not his “adjusted” Song score), the measure of occupational status.

**Econometric Specification.** Our analysis deviates somewhat from that of [Ward \(2023\)](#). To obtain a rank-rank regression model as in (10), we modify Ward’s regression in three ways. First, we run a rank-rank regression separately for each subpopulation of father-son pairs within which he computes ranks, i.e. within father-son birth cohort combinations. We remove father-son cohorts with fewer than 20 observations. Second, we do not use sample weights. Third,



**Figure 3:** Standard error of the indicated method divided by the correct standard error.

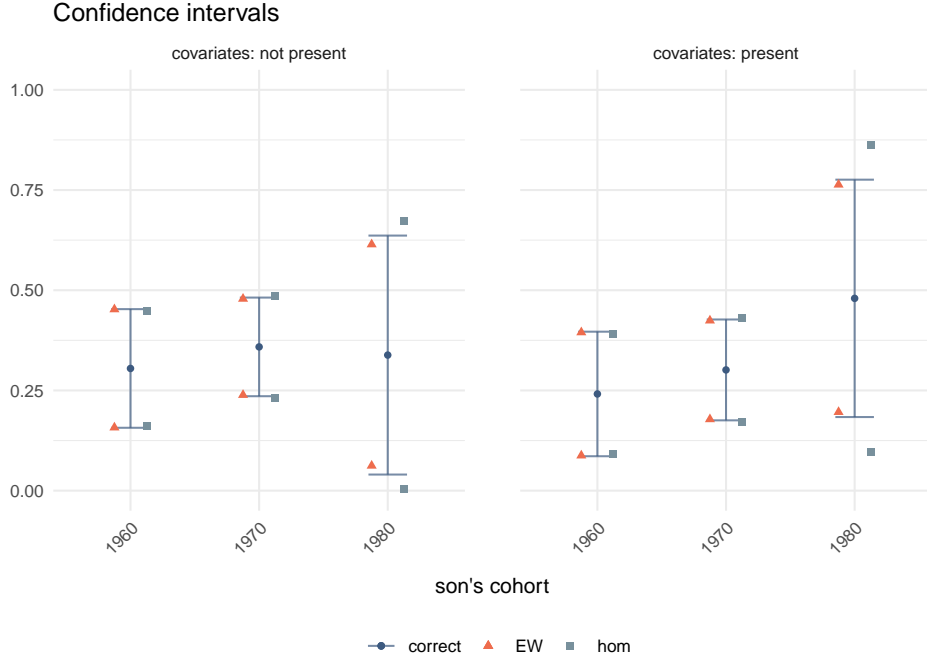
we do not average ranks over time. These three modifications to Ward (2023)’s regression specification allow us to interpret the regression slopes as rank-rank regression slopes and the model fits into (10). We estimate the rank-rank regression with and without the covariates used by Ward (2023). Since the occupational status measure is a discrete variable, the rank-rank slope we estimate is not a rank correlation (see Section 2.2). However, the occupational status measure has many support points so that different definitions of the rank do not substantially change estimates and standard errors. Therefore, we report results based on the largest ranks ( $\omega = 1$ ). Results for the smallest ranks ( $\omega = 0$ ) are similar.

**Results.** For the different father-son cohorts, Figure 3 shows the ratios of the hom and EW standard errors divided by the correct standard error. A ratio above (below) one means that the hom or EW standard error is larger (smaller) than the correct one. The left (right) panel reports ratios for the rank-rank regression without (with) covariates.

Some of the hom and EW standard errors are too small relative to the correct one, many are close to the correct one, but a few are substantially larger than the correct one. For instance, for the 1940-1980 father-son cohort, the hom standard error is up to 30% larger than the correct one.

The differences in standard errors translate into differences in confidence intervals for the rank-rank slope. Figure 4 shows 95% marginal confidence intervals for the rank-rank slope using data on father-son pairs for which the father is in the 1940 cohort. Most hom and EW confidence intervals are similar to the correct confidence interval, but for the son’s cohort of 1980 we see some differences: the hom confidence sets are longer than the correct ones and lead to somewhat conservative inference.

All confidence intervals are fairly wide, not permitting to draw tight conclusions about the mobility within father-son cohort pairs.



**Figure 4:** 95% marginal confidence intervals for the rank-rank slope for fathers in the 1940 cohort.

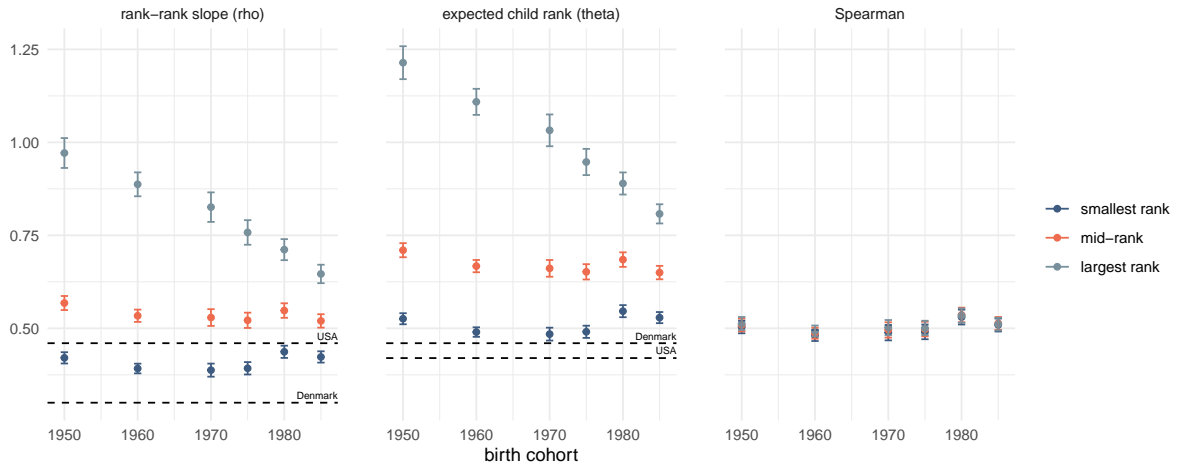
## 5.2 Intergenerational Mobility in India – Education

Asher et al. (2024) compute rank-based measures of mobility from data on children’s and parents’ years of education in India. Because education is observed on discrete points of support, they partially identify rank-rank slopes for a latent, continuously distributed measure of education. We complement their analysis by directly estimating rank-rank slopes from the observed, discrete education data, highlighting the sensitivity of the rank-rank slope to the way in which ties are handled, and by providing valid confidence intervals.

**Data.** We use the 2012 India Human Development Survey (IHDS) dataset from Asher et al. (2024) and focus on the relationship between fathers’ and sons’ years of education.

**Econometric Specification.** For each birth cohort of the children, we estimate the rank-rank slopes in (10) with  $W$  containing only a constant. Unlike Asher et al. (2024) we do not use sample weights. Ranks are defined as in (9) with  $\omega \in \{0, 0.5, 1\}$ .

**Results.** Fathers’ and sons’ education is observed on a small number of support points from 0 to 14; see Figure 15 in the Appendix. Because of this discreteness, we compare rank-rank slope estimates with different ways of handling ties in the ranks. For each birth cohort and each definition of the rank, Figure 5 shows estimates of the rank-rank slope  $\rho$ , of the expected child rank  $\theta_{0.25}$  as defined in Section 3.1, and of Spearman’s rank correlation.



**Figure 5:** Estimates of the rank-rank slope  $\rho$ , expected child rank  $\theta$ , and Spearman’s rank correlation by birth cohort and with 95% marginal confidence intervals. The colors refer to different definitions of ranks: “smallest rank” ( $\omega = 0$ ), “mid-rank” ( $\omega = 0.5$ ), and “largest rank” ( $\omega = 1$ ). The dashed-lines indicate similar measures of mobility for the U.S. and Denmark as reported in Asher et al. (2024).

Consider the graph on the left, which shows estimates and confidence intervals for the rank-rank slope. The estimates are very sensitive to the way in which ties are handled, particularly in the earlier birth cohorts. For instance, using the largest rank ( $\omega = 1$ ) leads to a rank-rank slope close to one with a relatively wide confidence interval. On the other hand, using the smallest rank ( $\omega = 0$ ) leads to a rank-rank slope of 0.42. In the latest birth cohort (1985), the largest and smallest ranks lead to rank-rank slopes of 0.65 and 0.42, still a substantial difference. For the 1950 birth cohort, according to the estimates based on the smallest ranks, mobility in India is higher than in the U.S. (i.e., the rank-rank slope is smaller), while the estimates based on the largest ranks imply that India has extremely low mobility with a rank-rank slope about twice as large as that of the U.S.. Also, based on the largest ranks, mobility in India has increased (rank-rank slopes have decreased) across birth cohorts. According to the estimates based on the smallest ranks, mobility has remained constant at a high level (low rank-rank slope). It is worth emphasizing that, for a given birth cohort, the difference in estimates is due only to the way in which ties in the ranks are handled, i.e. which value of  $\omega$  is used. Otherwise, the estimator and the data used are identical.

The graph in the middle shows estimates and confidence intervals for the expected child rank conditional on the parents’ rank being equal to 0.25. The graph is qualitatively similar to that of the rank-rank slope, except that all estimates lie above those for Denmark and the U.S.. Therefore, according to this measure, mobility in India appears to be higher than in Denmark and the U.S., regardless of how ties are handled.

The graph on the right shows estimates and confidence intervals for Spearman’s rank correlation for different ways of handling ties. Interestingly, the estimates and their confidence intervals are almost insensitive to the way in which ties are handled. The estimates are close to 0.5 for

all birth cohorts and all values of  $\omega$ , and the confidence intervals are all very narrow. To see why the estimates of the rank-rank slopes and of the Spearman’s rank correlation substantially differ, recall that the rank-rank slope  $\hat{\rho}$  is equal to the Spearman’s rank correlation  $\hat{\rho}_S$  multiplied by the ratio of standard deviations of the ranks:

$$\hat{\rho} = \hat{\rho}_S \sqrt{\frac{\sum_{i=1}^n (R_i^Y - \bar{R}^Y)^2}{\sum_{i=1}^n (R_i^X - \bar{R}^X)^2}}.$$

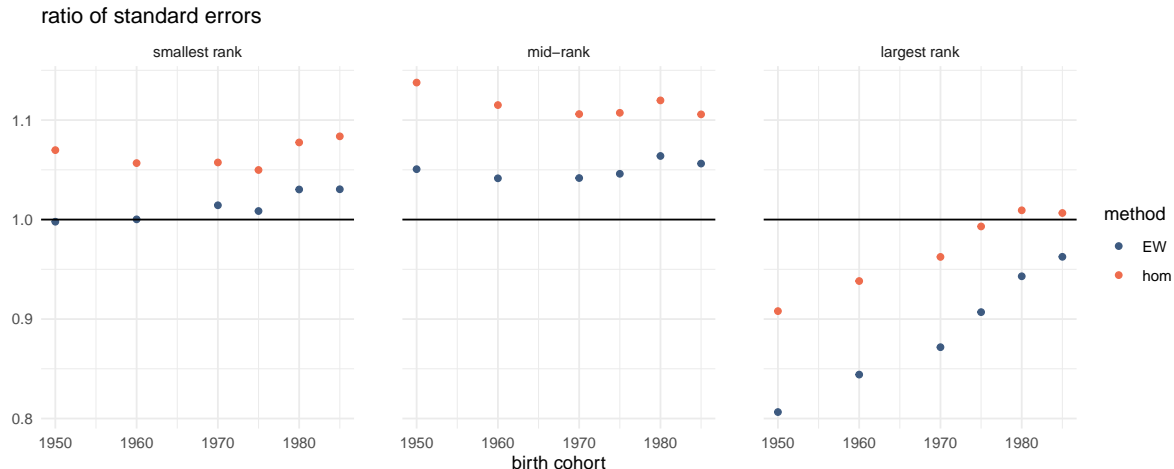
For continuous marginal distributions of  $X$  and  $Y$ , the ratio of standard deviations is equal to one, but when the marginal distributions are discrete, the ratio can take any value in  $(0, \infty)$ . While Spearman’s rank correlation is almost insensitive to how ties are handled and barely varies from birth cohort to birth cohort, the ratio of standard deviations is very sensitive to the way in which ties are handled and does vary substantially across birth cohorts. The ratios of standard errors is displayed in the right-most plot in Figure 2. When the smallest ranks ( $\omega = 0$ ) are employed, then the ratio of standard errors is close to 0.8 for all birth cohorts, while for the largest ranks ( $\omega = 1$ ), the ratio starts out at a value close to 1.9 for the 1950 birth cohort and then decreases towards about 1.25 for the 1985 birth cohort. Overall, the patterns in the ratios of standard errors mimics those found in the rank-rank slopes, while the Spearman’s rank correlation remains close to 0.5 for all birth cohorts and all definitions of the ranks. Therefore, the patterns seen in the rank-rank slopes across birth cohorts and their sensitivity to the way in which ties are handled originate from the variation in the ratios of standard errors rather than in Spearman’s rank correlation.

In this dataset, it seems possible to reach robust conclusions about the value of Spearman’s rank correlation, namely it being close to 0.5 for all birth cohorts. On the other hand, the rank-rank slope estimates are so sensitive to the definition of the ranks that they do not provide robust conclusions about whether the trend across birth cohorts is increasing or decreasing, nor whether the rank-rank slope in India is larger or smaller than that of the U.S..

Finally, Figure 6 shows the ratio of standard errors, i.e. the standard error of the indicated method (“hom” or “EW”) divided by the correct standard error. A ratio above (below) one means that the hom or EW standard error is larger (smaller) than the correct one. Both the hom and the EW standard errors may be too large or too small relative to the correct standard error, depending on how ties in the ranks are handled. For instance, for the smallest ranks, both standard errors are larger than the correct one for all birth cohorts with the EW being less conservative than the hom standard error. On the other hand, for the largest ranks, both standard errors are too small relative to the correct standard error. For the 1950 birth cohort, for instance, the EW standard error is about 20% too small and the hom standard error is about 10% too small.

### 5.3 Intergenerational Mobility in the U.S. – Income

In this section, we apply our new inference methods to the analysis of intergenerational mobility in U.S. commuting zones (CZs) as in [Chetty et al. \(2014, 2018\)](#); [Chetty and Hendren \(2018\)](#).



**Figure 6:** Standard error of the indicated method divided by the correct standard error. Each panel refers to a different definition of ranks: “smallest rank” ( $\omega = 0$ ), “mid-rank” ( $\omega = 0.5$ ), and “largest rank” ( $\omega = 1$ ).

Since the micro-data are not publicly available, we design a simulation experiment calibrated to the available information about commuting zones’ income distributions.

Suppose we are interested in learning about mobility in CZs of California. A local politician in a given CZ of the state may be interested in estimating mobility in their own CZ. In that case, she might run a rank-rank regression as in (10) using only data from that particular CZ. Importantly, ranks for children and parents are therefore computed in their respective distributions within that CZ. Second, a state-level politician may want to learn about mobility in all CZs of their state and perhaps compare mobility across CZs. In that case, she might want to run a rank-rank regression with clusters as in (15), where ranks are computed in the distribution of the whole state.

We apply our new inference methods to both types of regressions on simulated data, where the data-generating process has been calibrated to information from the CZs in California, and compare the results to methods using the homoskedastic and Eicker-White variance estimators. Subsequently, we show that the results from California are not special in the sense that similar results are found in other U.S. states. Finally, we explore the impact of top- and bottom-censoring of incomes on the estimators.

In both types of regressions, the CZ-specific regressions of the form (10) and the regression with clusters as in (15),  $W$  contains only a constant. Both regressions produce rank-rank slopes  $\rho_g$  for each CZ  $g$ . In addition, we compute the expected income rank of a child with parents at the  $p$ -th rank,  $\theta_{g,p}$  as defined in (19).

**Data.** From Chetty et al. (2022b), we obtain information about the CZ-specific marginal income distributions for 709 CZs.<sup>8</sup> Specifically, we obtain the CZ-specific mean and median for

<sup>8</sup>The information on the marginal income distributions is missing for some CZs, so we do not analyze all 741 CZs contained in the dataset by Chetty et al. (2022a).

2011-12 for child family income of children in the 1980-82 birth cohorts and the CZ-specific mean and median for 1996-2000 for the parent family income. In addition, we obtain CZ population sizes from the 2000 census. From [Chetty et al. \(2022a\)](#), we obtain estimates of  $\theta_{g,p}$  for two values  $p = 0.25$  and  $p = 0.75$  for each CZ together with their standard errors.

**Calibration.** First, we calibrate a CZ-specific copula for children’s and parents’ income. To this end, for each CZ  $g$ , we use the two estimates of  $\theta_{g,p}$  for  $p = 0.25$  and  $p = 0.75$  to solve for the implied estimate  $\hat{\rho}_g$  of the rank-rank slope  $\rho_g$ . We then consider two copula families, the normal and the t copulas. For each family we find the copula that matches the observed rank correlation  $\hat{\rho}_g$ .

Second, we calibrate CZ-specific marginal distributions for children’s and parents’ income. For each CZ  $g$  we fit a log-normal distribution that matches the observed CZ-specific mean and median incomes for children. Similarly, we fit a log-normal distribution that matches the observed CZ-specific mean and median incomes for parents.

Finally, we calibrate CZ-specific sample sizes proportional to the observed CZ population sizes.<sup>9</sup>

**Simulation Exercise.** We draw 1,000 Monte Carlo samples for each CZ. On each Monte Carlo sample, we estimate the two parameters of interest,  $\rho_g$  and  $\theta_{g,0.25}$ . For each, we compute three types of standard errors: the correct standard errors from [Theorem 5](#) or [Theorem 6](#), respectively, the OLS standard error assuming homoskedasticity of the regression errors (“hom”), and the Eicker-White (“EW”) standard error. In the rank-rank regression with clusters, the hom and EW standard errors for  $\theta_g, \rho_g$  are computed based on the residuals from CZ  $g$ .<sup>10</sup> Ranks are defined as in [\(9\)](#) with  $\omega = 1$ .

### 5.3.1 Mobility Across CZs in a Single State

We start by performing inference about mobility in the CZs of a single state, California. As indicated above, a local politician may be interested in estimating mobility in their own CZ and, thus, may run a rank-rank regression as in [\(10\)](#) using only data from that particular CZ. Similarly, a state-level politician may want to learn about mobility in all CZs of their state and perhaps compare mobility across CZs. In that case, she might want to run a rank-rank regression with clusters as in [\(15\)](#).

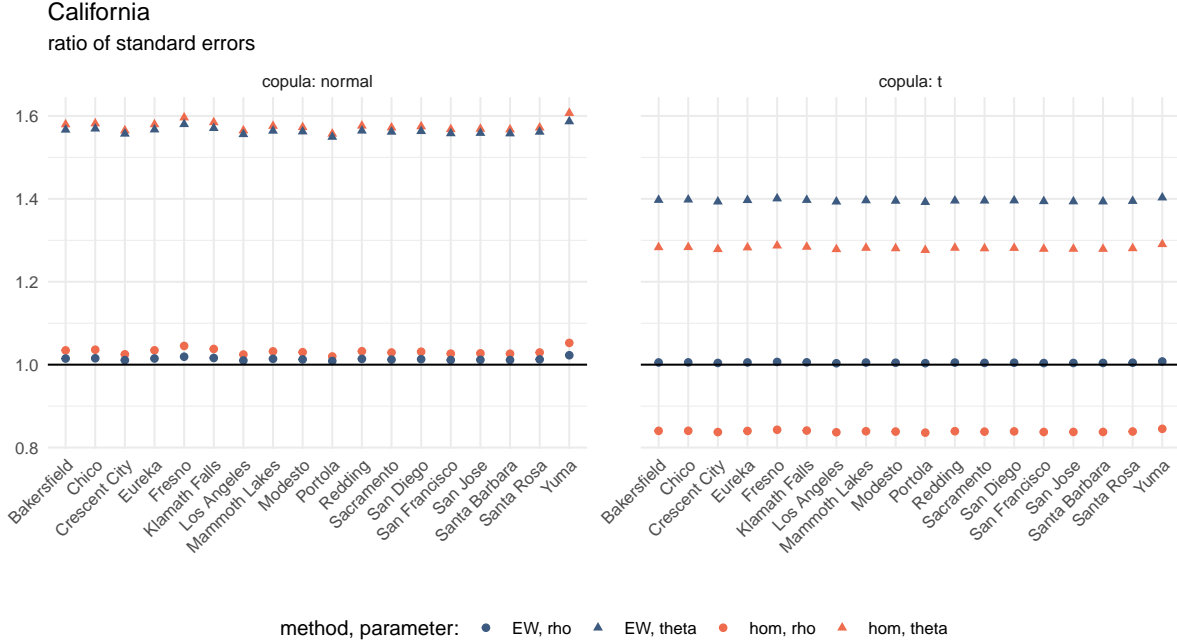
[Figures 7](#) and [8](#) show results for these two types of regressions. They plot the ratios of the hom and EW standard errors divided by the correct standard error. A ratio above (below) one

---

<sup>9</sup>We define a minimum ( $\underline{n} := 2,000$ ) and maximum ( $\bar{n} := 10,000$ ) sample size. Among all CZs in the U.S., the one with the smallest (largest) populations size is assigned  $\underline{n}$  ( $\bar{n}$ ). All other CZs are assigned a sample size between these two, proportional to their population size.

<sup>10</sup>An alternative approach would be to pool the residuals from all CZs to estimate the residual variance, similar to the regression in [\(15\)](#). The results from this approach are similar to those without pooling residuals and are not shown here.





**Figure 7:** Separate rank-rank regression within each CZ: standard error of the indicated method divided by the correct standard error.

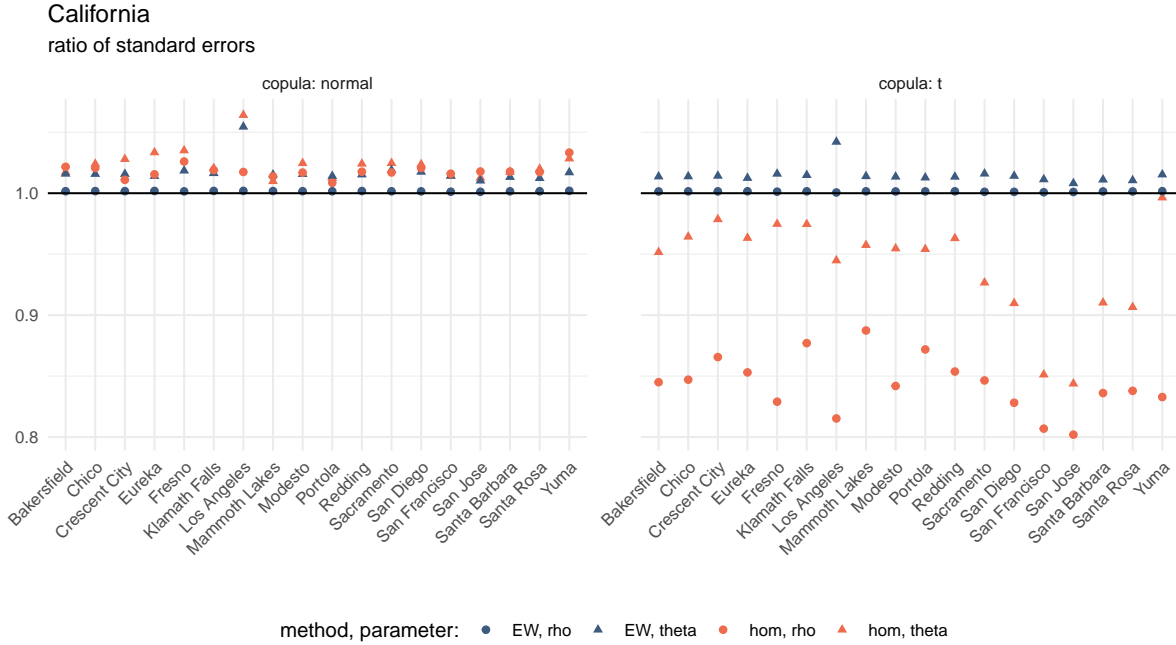
means that the hom or EW standard error is larger (smaller) than the correct one.

For the normal copula, the hom and EW standard errors are larger than the correct one in all CZs. In the separate rank-rank regressions, for the rank-rank slope  $\rho_g$ , the ratio of standard errors is close to one, but for the expected income rank  $\theta_{g,0.25}$ , the ratios are close to 1.6, i.e. the hom and EW standard errors are about 60% too large. For the rank-rank regression with clusters, the ratios are all fairly close to one except for  $\theta_{g,0.25}$  in Los Angeles, where the hom and EW standard errors are slightly over 5% too large.

For the t copula, on the other hand, the hom standard error for the rank-rank slope may be substantially (around 20%) smaller than the correct one. The hom standard error for the expected income rank may be around 30% too large in the separate rank-rank regressions, but up to 10% too small in the rank-rank regressions with clusters. The EW standard error may also be substantially too large (about 40% too large in the separate regressions), but is never too small.

The differences in standard errors translate into differences in confidence intervals. Figures 9 and 10 show 95% marginal confidence intervals for the mobility estimates based on the normal and the t copulas. Figures 16 and 17 in the Appendix show the corresponding figures for the rank-rank regression with clusters.

For the normal copula, confidence intervals for the rank-rank slope are similar across the three methods. For the expected income rank, however, the confidence intervals based on hom and EW variance estimators are substantially longer than those based on the correct variance estimator, and thus lead to conservative inference. The difference in length of the hom and EW

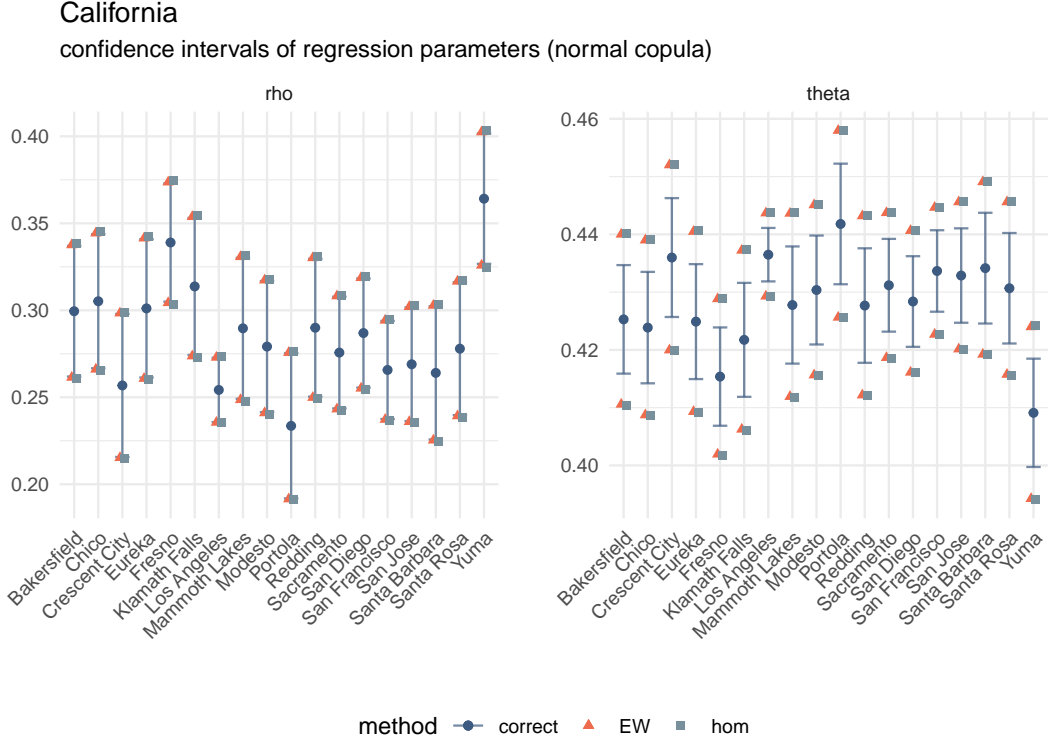


**Figure 8:** Rank-rank regression with clusters: standard error of the indicated method divided by the correct standard error.

compared to the correct confidence intervals is economically meaningful in the sense that the hom and EW confidence intervals include meaningful values of mobility that the correct intervals exclude. For instance, the hom and EW confidence intervals for Modesto, San Jose, Santa Rosa, and Santa Barbara all include Italy’s (Acciari et al. (2022) report a value of 0.445) and Canada’s (Corak (2020) reports a value of 0.444) values of mobility, but the correct intervals are shorter and exclude Italy’s and Canada’s values. Similarly, Crescent City’s hom and EW confidence intervals include Australia’s value of mobility (Deutscher and Mazumder (2023) report a value of 0.45), but the correct interval is shorter and does not include Australia’s value.

For the t copula, the hom confidence intervals for the rank-rank slope are too small while the EW confidence intervals are almost identical to the correct ones. The difference in length of the hom and the correct confidence intervals is economically meaningful in the sense that the hom confidence intervals exclude meaningful values of mobility that the correct intervals include. For instance, the hom confidence intervals for Redding and Mammoth Lakes imply that mobility in these two CZs is significantly different from that in Germany (Bratberg et al. (2017) report a rank-rank slope of 0.245) and Canada (Corak (2020) report a rank-rank slope of 0.242) while the correct confidence intervals are longer and include Germany’s and Canada’s values of mobility. Similarly, Santa Barbara’s hom confidence interval implies that mobility in this CZ is significantly different from that in Norway (Bratberg et al. (2017) report a rank-rank slope of 0.223) and Italy (Acciari et al. (2022) report a rank-rank slope of 0.220), but the correct confidence interval includes Norway’s and Italy’s values of mobility.

Comparisons of mobility across countries are common in the literature (Deutscher and



**Figure 9:** Separate rank-rank regression within each CZ: 95% marginal confidence intervals for the normal copula.

Mazumder (2023)) and the above discussion shows that our new confidence intervals may lead to different conclusions than those based on commonly used variance estimators.

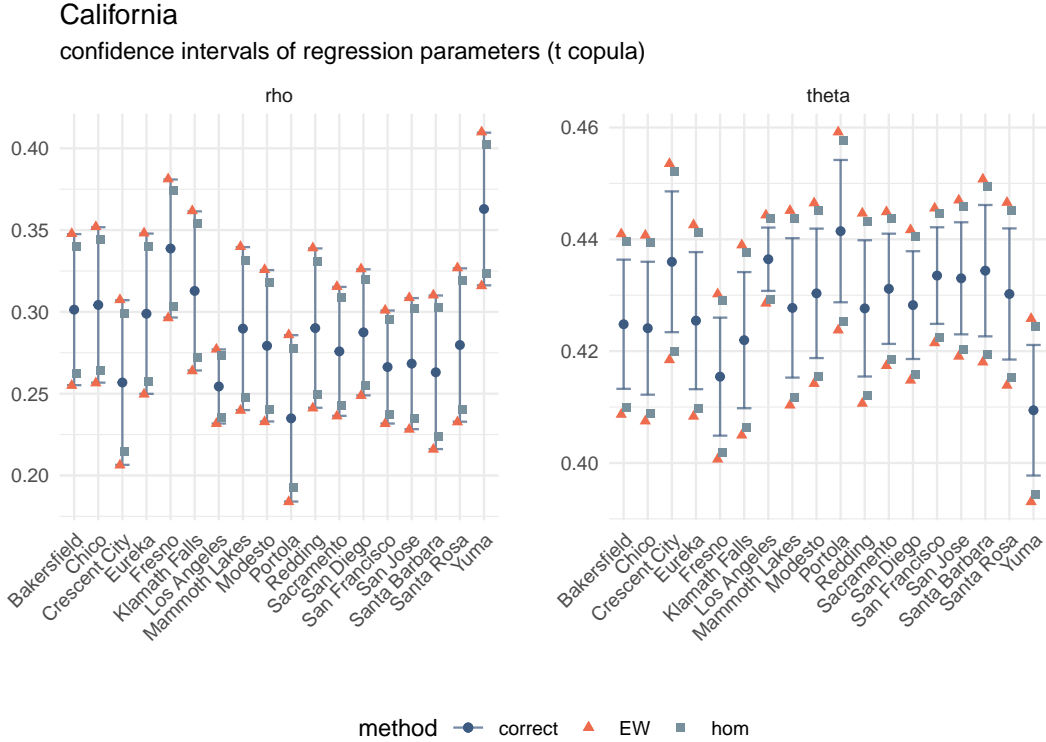
### 5.3.2 Mobility Across CZs in All States

We now turn to the 709 CZs across all U.S. states. For the separate rank-rank regressions, Figure 11 shows the histogram of ratios of standard errors across all CZs. Figure 18 in the Appendix shows the corresponding figure for the rank-rank regression with clusters. Both figures illustrate that the results for California are not special. In particular, as in California, most ratios for the normal copula and the expected income rank parameter take values around 1.6, so that most hom and EW standard errors are about 60% too large. Similarly, as in California, most hom standard errors for the t copula and the rank-rank slope are about 20% too small.

### 5.3.3 The Impact of Censoring

In Section 2.2 and Remark 2, we have already argued that pointmasses in the marginal distributions of income create ties in the ranks and that the way in which ties are handled has an impact on the estimand and the statistical properties of the OLS estimator.

First, we study the impact of top- and bottom-censoring of the incomes on the estimand  $\rho$ . To this end we draw a large sample (of size  $10^6$ ) from the (continuous) calibrated income



**Figure 10:** Separate rank-rank regression within each CZ: 95% marginal confidence intervals for the t copula.

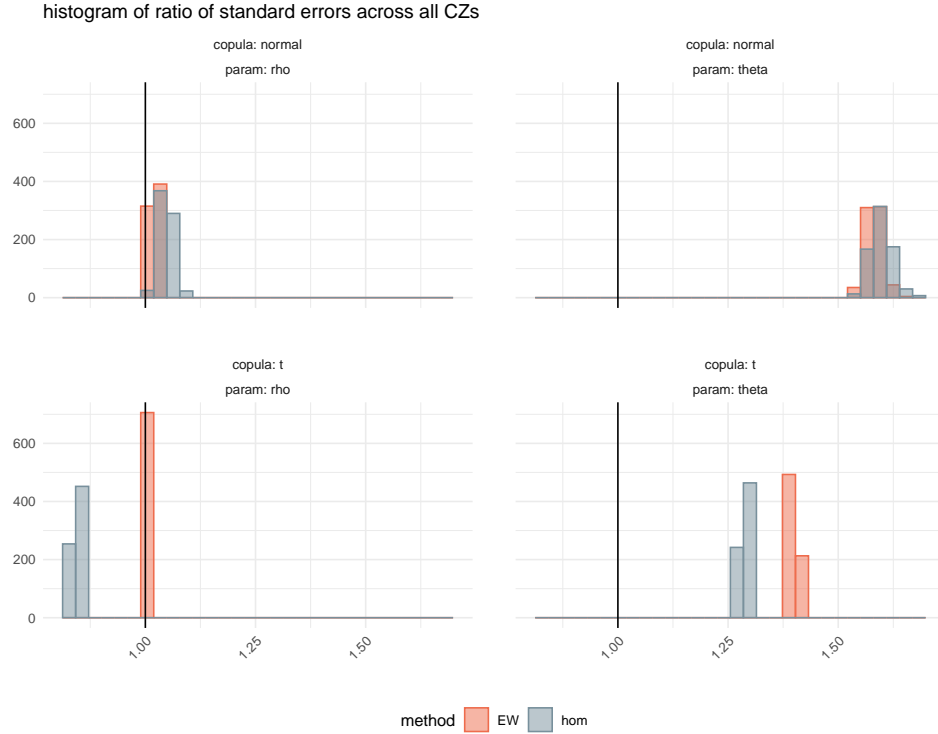
distribution of the CZ Los Angeles using the Gaussian copula. Then, parental incomes are censored below (“bottom censoring”) or above (“top censoring”) at the  $q$ -th or the  $(1 - q)$ -th quantile. We successively increasing  $q$  from zero to 0.25.

Figure 12 shows how the rank-rank slope and Spearman’s rank correlation change with the amount of censoring  $q$  and with the way in which ties are handled.

Consider the first row of the figure. At  $q = 0$  there is no censoring and the joint distribution of incomes is continuous, so the rank correlation and the rank-rank slope are equal. With more censoring (larger  $q$ ), the joint distribution features larger pointmasses and the two estimands diverge. Without censoring, the rank-rank slope is equal to 0.254 while with 25% top-censoring, the rank-rank slope takes a value of 0.287. Among the ranking of all 709 commuting zones (CZs) that we analyze, this change in mobility corresponds to Los Angeles moving from 59th place to 134th place. This is a substantial change in mobility. Of course, 25% of the sample being censored is an extreme case, but we see significant sensitivity of the rank-rank slope with respect to  $q$  even for smaller values. In addition, discreteness stemming from sources other than censoring may have similarly substantial impacts on the estimand.

While for top-censoring, the rank-rank slope increases with  $q$ , for bottom-censoring it decreases with  $q$ .

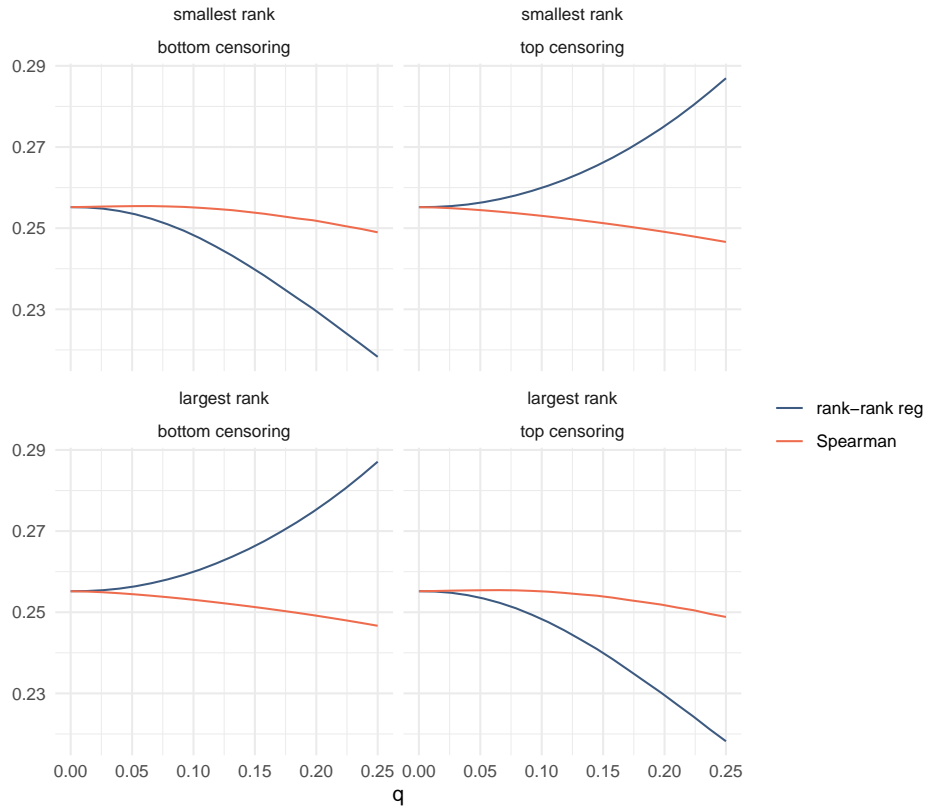
The first row of the figure shows the estimands for  $\omega = 0$ , i.e. the smallest rank. The second row shows the same estimands when  $\omega = 1$ , i.e. largest ranks. The rank correlation is almost



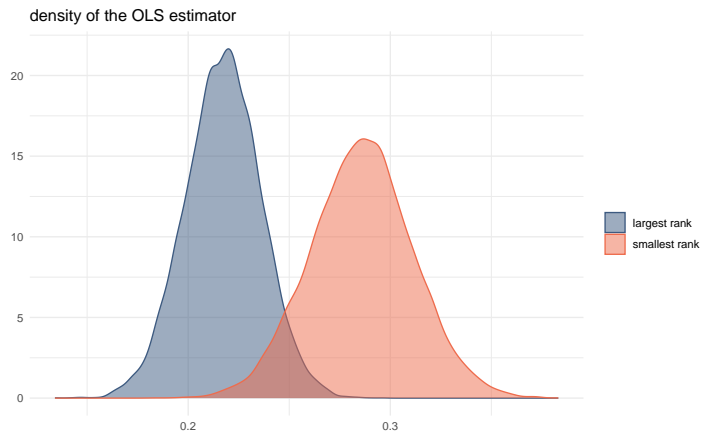
**Figure 11:** Separate rank-rank regression within each CZ: histograms of ratios of standard errors.

the same for the two definitions of the rank, but the rank-rank slope substantially changes. Handling ties differently ( $\omega = 1$ ) than in the first row of the figure ( $\omega = 0$ ) means that Los Angeles' mobility measure drops from 0.254 without censoring to 0.218 with 25% top-censoring. This change in mobility corresponds to Los Angeles moving from 59th place up to 20th place in the ranking of all 709 CZs.

Second, the way in which ties are handled also impacts the statistical properties of the OLS estimator of the rank-rank slope. Figure 13 illustrates this point by showing the estimated distribution of the estimator for the two different definitions of the ranks for  $\omega = 0$  and  $\omega = 1$ . To create this graph, we simulated 10,000 samples of size 2,000 from the income distribution calibrated to data from Los Angeles with 25% top-coding and the Gaussian copula. In Figure 12, we have already seen how the estimand changes depending on whether  $\omega = 0$  or  $\omega = 1$ . In Figure 13, we see this reflected in the different means of the two distributions of the estimator. In addition, however, the variance of the two estimators is also different. If ties are handled by setting  $\omega = 0$ , the variance is larger compared to the estimator's variance when  $\omega = 1$ .



**Figure 12:** Spearman’s rank correlation (“Spearman”) and the slope coefficient from a rank-rank regression (“rank-rank reg”), where  $Y$  is a child’s income and  $X$  their parent’s income. The graphs in the first column (“bottom censoring”) show the two estimands when all parental incomes up to the  $q$ -th quantile are set to zero, while those in the second column (“top censoring”) show the same quantities when all parental incomes above the  $(1 - q)$ -th quantile are set equal to the income at that quantile. The first row employs the smallest rank ( $\omega = 0$ ) and the second row the largest rank ( $\omega = 1$ ).



**Figure 13:** Estimated density of the OLS estimator of the rank-rank slope with two different ways of handling ties: either employing the smallest rank ( $\omega = 0$ ) or the largest rank ( $\omega = 1$ ). The densities are estimated based on simulated data from the income distribution calibrated to data from Los Angeles with 25% top-censoring.

## References

- ABRAMITZKY, R., L. BOUSTAN, E. JACOME, AND S. PEREZ (2021): “Intergenerational Mobility of Immigrants in the United States over Two Centuries,” *American Economic Review*, 111, 580–608.
- ACCIARI, P., A. POLO, AND G. L. VIOLANTE (2022): “And Yet It Moves: Intergenerational Mobility in Italy,” *American Economic Journal: Applied Economics*, 14, 118–63.
- ASHER, S., P. NOVOSAD, AND C. RAFKIN (2024): “Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups,” *American Economic Journal: Applied Economics*, 16, 66–98.
- BAGDE, S., D. EPPLE, AND L. TAYLOR (2016): “Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India,” *American Economic Review*, 106, 1495–1521.
- BAZYLK, S., M. MOGSTAD, J. P. ROMANO, A. M. SHAIKH, AND D. WILHELM (2024): “Finite- and large-sample inference for ranks using multinomial data with an application to ranking political parties,” Technical report.
- BICKEL, P., AND D. FREEDMAN (1981): “Some asymptotic theory for the bootstrap,” *Annals of Statistics*, 9, 1196–1217.
- BLACK, S. E., P. J. DEVEREUX, P. LUNDBORG, AND K. MAJLESI (2019): “Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours,” *The Review of Economic Studies*, 87, 1683–1725.
- BORKOWF, C. B. (2002): “Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman’s rank correlation,” *Computational Statistics & Data Analysis*, 39, 271–286.
- BRATBERG, E., J. DAVIS, B. MAZUMDER, M. NYBOM, D. D. SCHNITZLEIN, AND K. VAAGE (2017): “A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US,” *The Scandinavian Journal of Economics*, 119, 72–101.
- CAGÉ, J., N. HERVÉ, AND M.-L. VIAUD (2019): “The Production of Information in an Online World,” *The Review of Economic Studies*, 87, 2126–2164.
- CHATTERJEE, S. (2023): “A survey of some recent developments in measures of association,” *arXiv:2211.04702*.
- CHERNOZHUKOV, V., J. ESCANCIANO, H. ICHIMURA, W. NEWEY, AND J. ROBINS (2022): “Locally robust semiparametric estimation,” *Econometrica*, 90, 1501–1535.

- CHETTY, R., J. FRIEDMAN, N. HENDREN, M. R. JONES, AND S. R. PORTER (2022a): “Replication Data for: The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility,” <https://doi.org/10.7910/DVN/NKCQM1>.
- CHETTY, R., J. N. FRIEDMAN, N. HENDREN, M. R. JONES, AND S. R. PORTER (2018): “The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility,” Working Paper 25147, NBER.
- CHETTY, R., J. N. FRIEDMAN, N. HILGER, E. SAEZ, D. W. SCHANZENBACH, AND D. YAGAN (2011): “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star,” *The Quarterly Journal of Economics*, 126, 1593–1660.
- CHETTY, R., AND N. HENDREN (2018): “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects,” *The Quarterly Journal of Economics*, 133, 1107–1162.
- CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *The Quarterly Journal of Economics*, 129, 1553.
- (2022b): “Replication Data for: Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” <https://doi.org/10.7910/DVN/NALG3E>.
- CORAK, M. (2020): “The Canadian Geography of Intergenerational Income Mobility,” *The Economic Journal*, 130, 2134–2174.
- DE LA CROIX, D., AND M. GONI (2022): “Nepotism vs. Intergenerational Transmission of Human Capital in Academia (1088–1800),” discussion paper.
- DAHL, M., AND T. DELEIRE (2008): “The Association between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data,” Discussion Paper 1342-08, Institute for Research on Poverty.
- DEUTSCHER, N., AND B. MAZUMDER (2023): “Measuring Intergenerational Income Mobility: A Synthesis of Approaches,” *Journal of Economic Literature*, 61, 988–1036.
- DUDLEY, R. M. (2002): *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics: Cambridge University Press, 2nd edition.
- (2014): *Uniform Central Limit Theorems*: Cambridge University Press, Cambridge.
- FAGERENG, A., M. MOGSTAD, AND M. RØNNING (2021): “Why Do Wealthy Parents Have Wealthy Children?” *Journal of Political Economy*, 129, 703–756.
- FAIA, E., AND V. PEZONE (2023): “The Cost of Wage Rigidity,” *The Review of Economic Studies*, 91, 301–339.



- FANG, Z., AND A. SANTOS (2019): “Inference on directionally differentiable functions,” *Review of Economic Studies*, 86, 377–412.
- GENEST, C., J. NEŠLEHOVA, AND B. RÉMILLARD (2013): “On the estimation of Spearman’s rho and related tests of independence for possibly discontinuous multivariate data,” *Journal of Multivariate Analysis*, 98, 214–228.
- GHOSH, A., S. I. M. HWANG, AND M. SQUIRES (2023): “Economic Consequences of Kinship: Evidence From U.S. Bans on Cousin Marriage,” *The Quarterly Journal of Economics*, 138, 2559–2606.
- GRÖNQVIST, H., J. P. NILSSON, AND P.-O. ROBLING (2020): “Understanding How Low Levels of Early Lead Exposure Affect Children’s Life Trajectories,” *Journal of Political Economy*, 128, 3376–3433.
- HOEFFDING, W. (1948): “A Class of Statistics with Asymptotically Normal Distribution,” *The Annals of Mathematical Statistics*, 19, 293 – 325.
- HUFFMAN, D., C. RAYMOND, AND J. SHVETS (2022): “Persistent Overconfidence and Biased Memory: Evidence from Managers,” *American Economic Review*, 112, 3141–75.
- KLEIN, M., T. WRIGHT, AND J. WIECZOREK (2020): “A joint confidence region for an overall ranking of populations,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69, 589–606.
- LEE, S., AND J. LIN (2017): “Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income,” *The Review of Economic Studies*, 85, 663–694.
- MESFIOUI, M., AND J.-F. QUESSY (2010): “Concordance measures for multivariate non-continuous random vectors,” *Journal of Multivariate Analysis*, 101, 2398–2410.
- MOGSTAD, M., J. P. ROMANO, A. M. SHAIKH, AND D. WILHELM (2024): “Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries,” *The Review of Economic Studies*, 91, 476–518.
- MOGSTAD, M., AND G. TORSVIK (2023): “Family background, neighborhoods, and intergenerational mobility,” in *Handbook of the Economics of the Family, Volume 1* ed. by Lundberg, S., and Voena, A. Volume 1 of Handbook of the Economics of the Family: North-Holland, Chap. 6, 327–387.
- MURPHY, R., AND F. WEINHARDT (2020): “Top of the Class: The Importance of Ordinal Rank,” *The Review of Economic Studies*, 87, 2777–2826.
- NEŠLEHOVA, J. (2007): “On rank correlation measures for non-continuous random variables,” *Journal of Multivariate Analysis*, 98, 544–567.

- NEWKEY, W. (1994): “The asymptotic variance of semiparametric estimators,” *Econometrica*, 62, 1349–1382.
- NYBOM, M., AND J. STUHLER (2023): “Interpreting Trends in Intergenerational Mobility,” *Journal of Political Economy*, forthcoming.
- OLIVETTI, C., AND M. D. PASERMAN (2015): “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940,” *American Economic Review*, 105, 2695–2724.
- ORNSTEIN, P., AND J. LYHAGEN (2016): “Asymptotic properties of Spearman’s rank correlation for variables with finite support,” *PLOS one*, 11, 1–7.
- SERFLING, R. J. (2002): *Approximation Theorems of Mathematical Statistics*: John Wiley & Sons, Inc.
- SHARKEY, P., AND G. TORRATS-ESPINOSA (2017): “The effect of violent crime on economic mobility,” *Journal of Urban Economics*, 102, 22–33.
- SONG, X., C. G. MASSEY, K. A. ROLF, J. P. FERRIE, J. L. ROTHBAUM, AND Y. XIE (2020): “Long-term decline in intergenerational mobility in the United States since the 1850s,” *Proceedings of the National Academy of Sciences*, 117, 251–258.
- WARD, Z. (2022): “Internal Migration, Education, and Intergenerational Mobility,” *Journal of Human Resources*, 57, 1981–2011.
- (2023): “Intergenerational Mobility in American History: Accounting for Race and Measurement Error,” *American Economic Review*, 113, 3213–48.

# Appendix

## A The Survey on the Use of Regressions Involving Ranks

We searched for papers on Google Scholar. For instance, to find articles in the Review of Economic Studies, we used the search phrase “rank regression source:”Review of Economic Studies”” and then restricted the years of publication to 2013 to present (February 2024 at the time of the search). For the American Economic Review, we excluded Papers & Proceedings, AER Insights, and comments. For the Journal of Political Economy, we excluded JPE Micro and JPE Macro.

We then manually went through all search results, looking for the occurrence of the word “rank” in each paper. Many papers were immediately discarded if, for instance the word only occurred in the context of “rank conditions” or in papers without an empirical analysis. For each paper, we made a judgement about whether the regressor, the outcome variable, or both in a regression had been transformed into ranks, classifying them as “rank-rank”, “level-rank”, or “rank-level”. For papers containing at least one of these types of regressions, we then classified the type of estimator being used into the categories “OLS” (containing papers that explicitly state that OLS was used and those that didn’t explicitly state the method used, but we judged to most likely have been OLS<sup>11</sup>), “TSLS”, “nonparametric” (e.g., kernel regression, local-polynomial regression, bin-scatter), and “other” (e.g., probit, logit, regression discontinuity designs). Next, we classified the methods used for computation of standard errors into the categories “homoskedastic” (papers explicitly stating that they used the textbook homoskedastic error formula), “robust” (any estimator referred to as “robust” or “heteroskedasticity-robust”, but does not contain the word “cluster”), “clustered” (any method whose description contains the words based on the stem “cluster”), “unknown” (for cases in which we could not find any description of the method used), “none” (for cases in which no standard errors were computed), and “other” (e.g., bootstrap, permutation inference).

## B Differences in Asymptotic Variances

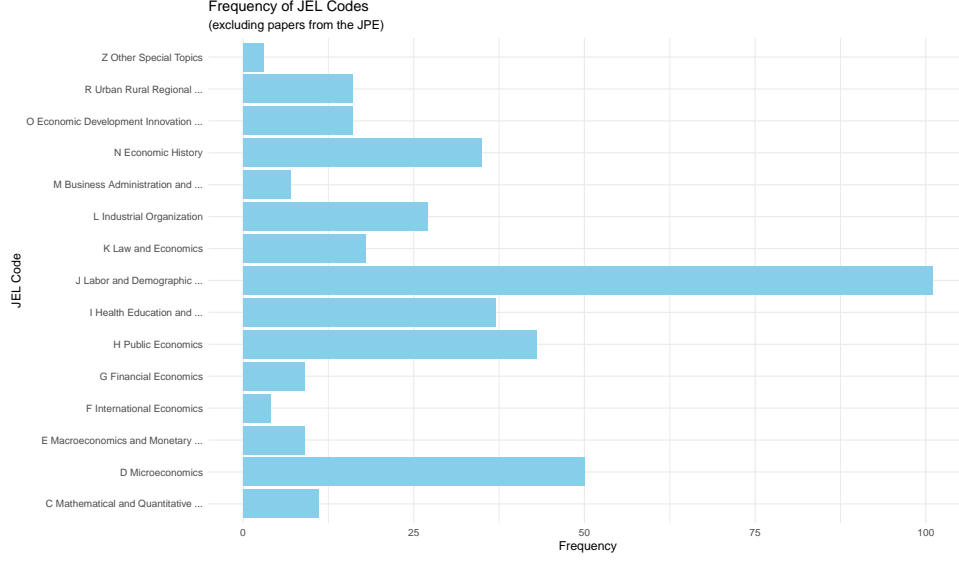
### B.1 Asymptotic Variances as Functionals of the Copula

Under the assumption that  $F$  is continuous, all three asymptotic variances analyzed in Section 2 can be written as functionals of the copula  $C$ . First, one can use results in [Borkowf \(2002\)](#) to show that

$$\sigma^2 = 144 \{-9\theta_1^2 + \theta_2 + 2\theta_3 + 2\theta_4 + 2\theta_5\} \quad (25)$$

---

<sup>11</sup>These are cases in which the paper contains a description of the form “. . . was regressed on . . .” without any further details.



**Figure 14:** Number of (paper  $\times$  regression type  $\times$  standard error type) combinations by top-level JEL-code. The sample consists of papers published in American Economic Review, Quarterly Journal of Economics, and Review of Economic Studies between January 2013 and February 2024. The Journal of Political Economy is excluded because it does not publish JEL codes for its articles.

where

$$\begin{aligned} \theta_1 &= \int_0^1 \int_0^1 uv dC(u, v) \\ \theta_2 &= \int_0^1 \int_0^1 (1-u)^2(1-v)^2 dC(u, v) \\ \theta_3 &= \int_0^1 \int_0^1 \int_0^1 \int_0^1 (1-u-\bar{v}+C(u, \bar{v}))(1-\bar{u})(1-v) dC(u, v) dC(\bar{u}, \bar{v}) \\ \theta_4 &= \int_0^1 \int_0^1 \int_0^1 \int_0^1 \min\{1-u, 1-\bar{u}\}(1-v)(1-\bar{v}) dC(u, v) dC(\bar{u}, \bar{v}) \\ \theta_5 &= \int_0^1 \int_0^1 \int_0^1 \int_0^1 \min\{1-v, 1-\bar{v}\}(1-u)(1-\bar{u}) dC(u, v) dC(\bar{u}, \bar{v}) \end{aligned}$$

Second, since  $\rho = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3$ , the homoskedastic asymptotic variance of the OLS estimator can be written as

$$\sigma_{hom}^2 = 1 - \left( 12 \int_0^1 \int_0^1 C(u, v) dudv - 3 \right)^2. \quad (26)$$

Finally, noticing that

$$M_{kl} = \int_0^1 \int_0^1 (u - 1/2)^k (v - 1/2)^l dC(u, v),$$

it is clear that

$$\sigma_{EW}^2 = 144 \left( M_{22} - 2\rho M_{31} + \frac{\rho^2}{80} \right)$$

is a functional of  $C$ . Interestingly, all three variances depend on  $C$ , and only on  $C$ .

## B.2 Optimizing the Difference in Variances Within Families of Copulas

This section provides details on the calculations of the maximal and minimal differences of variances presented in Figure 1. Specifically, we want to minimize and maximize the difference  $\sigma_j^2 - \sigma^2$  for  $j \in \{hom, EW\}$ . As shown in Appendix B.1, both variances depend on the copula  $C$  of  $F$ , and only on  $C$ . To make this dependence explicit, we write  $\sigma^2(C)$  and  $\sigma_{hom}^2(C)$ . Consider the maximization problem first. We want to solve

$$\max_{C \in \mathcal{C}} \sigma_j^2(C) - \sigma^2(C), \quad (27)$$

where  $\mathcal{C}$  is a family of copulas. We consider three families, the Gaussian family, the (Student-)t family, and one with quadratic dependence.

Denote by  $G_\theta$  the bivariate Gaussian copula with correlation  $\theta$  (both variances are equal to one, means are equal to zero). The Gaussian family is then defined as

$$\mathcal{C}_{Gaussian} := \{G_\theta : \theta \in [0, 1]\}.$$

Let  $T_\theta$  denote a (Student-)t copula with one degree of freedom and dependence parameter  $\theta$ . Then, the t family we consider is

$$\mathcal{C}_t := \{T_\theta : \theta \in [0, 1]\}.$$

Finally, consider the following “quadratic” data-generating process.  $X \sim U[-1/2, 1/2]$ ,  $Y = 1/2 + \theta X + (1 - \theta)X^2 + \varepsilon$ , where  $\varepsilon \sim N(0, 10^{-6})$ . Denote by  $F_\theta$  the joint distribution of  $(X, Y)$  and by  $Q_\theta$  the copula of  $F_\theta$ . Then, the quadratic family we consider is defined as

$$\mathcal{C}_{quadratic} := \{Q_\theta : \theta \in [0, 1]\}.$$

We want to solve (27) with  $\mathcal{C}$  being replaced by each of the three families as well as the corresponding minimization problem. These are optimization problems in the scalar parameter  $\theta \in [0, 1]$ , which we solve by simulation combined with a grid search. First, we define an equidistant grid of  $[0, 1]$  of size 200. For each parameter value  $\theta$  on this grid, we then draw 10,000 samples of size 2,000 from the bivariate copula in  $\mathcal{C}$  with that parameter. Then, we compute the value of the objective function in (27) for each of these draws and then average over all Monte Carlo samples. Denote the maximizer  $\theta$  on the grid leading to the largest value of the objective function by  $\theta^*$ . The left graph of panel A in Figure 1 then plots the magnitude of  $\sigma^2(C_{\theta^*})$  (“correct”), of  $\sigma_{hom}^2(C_{\theta^*})$  (“hom”) and the rank correlation  $\rho$  of the copula  $C_{\theta^*}$  in the given parametric family.

The black dots are computed as follows. In the data from Section 5.3, Los Angeles has a rank correlation of 0.254. For each of the parametric families, we can then find the parameter value  $\theta_{LA} \in [0, 1]$  so that the copula from that parametric family has a rank correlation equal to 0.254. Besides the value of  $\rho$ , the dots then show the values of  $\sigma^2(C_{\theta_{LA}})$  and  $\sigma_j^2(C_{\theta_{LA}})$  for  $j \in \{hom, EW\}$ .

The right graph in panel A of Figure 1 plots the quantities for the copula that solves the minimization problem corresponding to (27). Panel B of the figure shows the corresponding graphs for the optimization problems with  $j = EW$ .

## C Additional Results for the Empirical Applications

### C.1 Intergenerational Mobility in India – Education

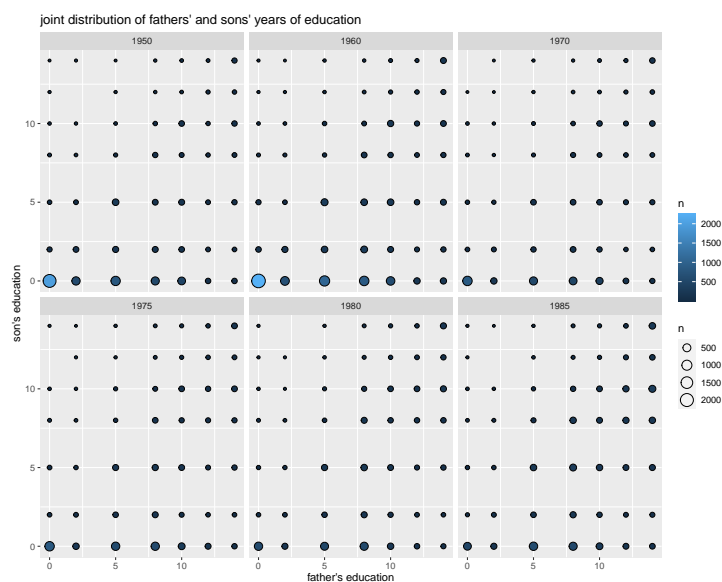


Figure 15: Count of observations for each pair of father and son years of education, by birth cohort of the son.

### C.2 Intergenerational Mobility in the U.S. – Income

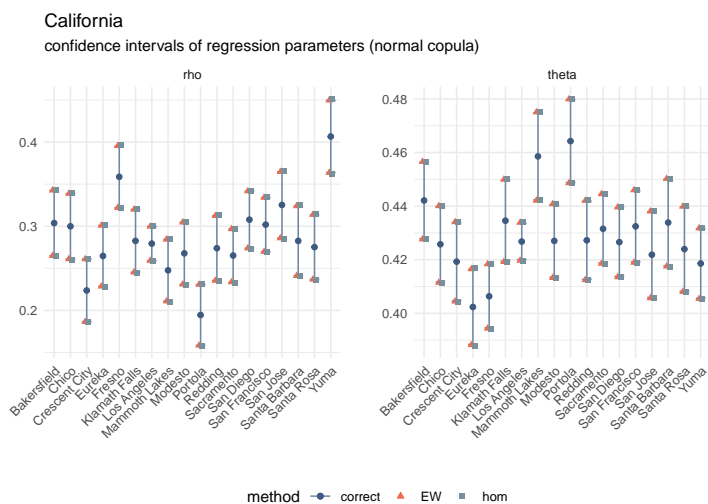


Figure 16: Rank-rank regression with clusters: 95% marginal confidence intervals for the normal copula.

California  
confidence intervals of regression parameters (t copula)

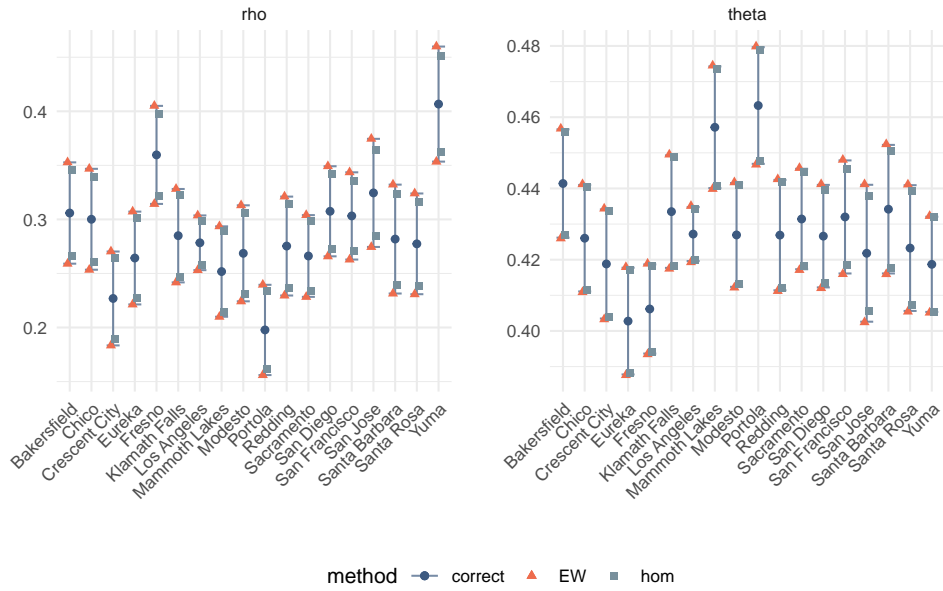


Figure 17: Rank-rank regression with clusters: 95% marginal confidence intervals for the t copula.

histogram of ratio of standard errors across all CZs

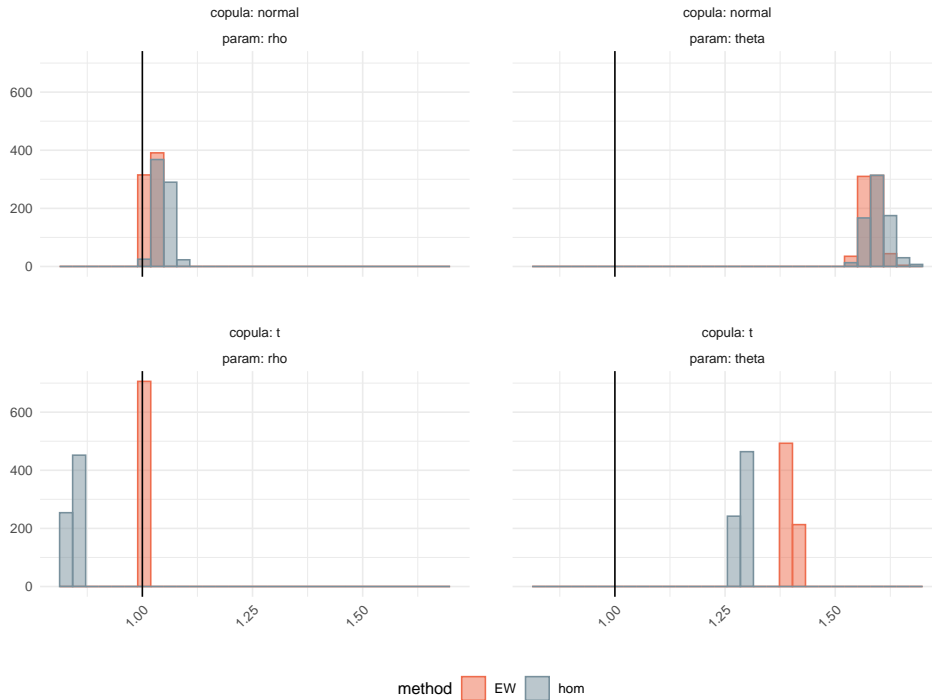


Figure 18: Rank-rank regression with clusters: histograms of ratios of standard errors.

# Online Appendix

## D Asymptotic Theory for All Regression Coefficients

### D.1 Rank-Rank Regressions

In this section, we present the asymptotic theory for all coefficients in the rank-rank regression model studied in Section 3:

$$R_Y(Y) = \rho R_X(X) + W'\beta + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \right] = 0, \quad (28)$$

where we had already introduced the equation projecting  $R_X(X)$  onto the other regressors in  $W$ :

$$R_X(X) = W'\gamma + \nu, \quad E[\nu W] = 0. \quad (29)$$

To study the asymptotic behavior of the coefficients  $\beta$ , we now also introduce some additional projections. Let  $W_l$  denote an element of  $W := (W_1, \dots, W_p)'$  and by  $W_{-l}$  the vector of all elements of  $W$  except the  $l$ -th. We now introduce the projection of  $W_l$  onto  $R_X(X)$  and the remaining regressors  $W_{-l}$ : for any,  $l = 1, \dots, p$ ,

$$W_l = \tau_l R_X(X) + W'_{-l} \delta_l + v_l, \quad E \left[ v_l \begin{pmatrix} R_X(X) \\ W_{-l} \end{pmatrix} \right] = 0, \quad (30)$$

where  $\tau_1, \dots, \tau_p$  are scalar constants and  $\delta_1, \dots, \delta_p$  are  $(p-1)$ -dimensional vectors of constants.

**Assumption 8.** For  $l = 1, \dots, p$ , the random variable  $v_l$  is such that  $E[v_l^2] > 0$ .

Consider the OLS estimator in (11):

$$\begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \left( \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} R_i^Y. \quad (31)$$

Let  $\sigma_\nu^2$ ,  $h_{0,1} := h_1$ ,  $h_{0,2} := h_2$ , and  $h_{0,3} := h_3$  be as defined in Theorem 1. For  $l = 1, \dots, p$ , further define  $\sigma_{v_l}^2 := E[v_l^2]$ ,

$$\begin{aligned} \phi_0(x, w, y) &:= \frac{1}{\sigma_\nu^2} [h_{0,1}(x, w, y) + h_{0,2}(x, y) + h_{0,3}(x)] \\ \phi_l(x, y, w) &:= \frac{1}{\sigma_{v_l}^2} [h_{l,1}(x, w, y) + h_{l,2}(x, y) + h_{l,3}(x)] \end{aligned}$$

and

$$\begin{aligned} h_{l,1}(x, w, y) &:= (R_Y(y) - \rho R_X(x) - w'\beta)(w_l - \tau_l R_X(x) - w'_{-l} \delta_l), \\ h_{l,2}(x, y) &:= E[(I(y, Y) - \rho I(x, X) - W'\beta)(W_l - \tau_l R_X(X) - W'_{-l} \delta_l)], \\ h_{l,3}(x) &:= E[(R_Y(Y) - \rho R_X(X) - W'\beta)(W_l - \tau_l I(x, X) - W'_{-l} \delta_l)]. \end{aligned}$$

Finally, let  $\psi_i := (\phi_0(X_i, W_i, Y_i), \phi_1(X_i, W_i, Y_i), \dots, \phi_p(X_i, W_i, Y_i))'$ .

With the additional assumption and notation, one can derive the following joint asymptotic normality result:



**Theorem 5.** Suppose that (28)–(31) hold and that Assumptions 1–3, and 8 are satisfied. Then

$$\sqrt{n} \begin{pmatrix} \widehat{\rho} - \rho \\ \widehat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma),$$

where  $\Sigma := E[\psi_i \psi_i']$  and  $\psi_i$ 's are as defined above in this section.

Theorem 1 has already shown the asymptotic variance of  $\widehat{\rho}$ . Theorem 5 adds to that result by providing the asymptotic variance for individual components of  $\widehat{\beta}$ : for all  $l = 1, \dots, p$ , we have

$$\sqrt{n}(\widehat{\beta}_l - \beta_l) \rightarrow_D N(0, \sigma_{\beta_l}^2),$$

where

$$\sigma_{\beta_l}^2 := \frac{1}{\sigma_{v_l}^4} E \left[ (h_{l,1}(X, W, Y) + h_{l,2}(X, Y) + h_{l,3}(X))^2 \right].$$

## D.2 Rank-Rank Regressions With Clusters

In this section, we present the asymptotic theory for all coefficients in the rank-rank regression model with clusters studied in Section 4.1:

$$R_Y(Y) = \sum_{g=1}^{n_G} \mathbb{1}\{G = g\} (\rho_g R_X(X) + W' \beta_g) + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \middle| G \right] = 0 \text{ a.s.}, \quad (32)$$

where we had already introduced the equation projecting  $R_X(X)$  onto the other regressors in  $W$ :

$$R_X(X) = \sum_{g=1}^{n_G} \mathbb{1}\{G = g\} W' \gamma_g + \nu, \quad E[\nu W | G] = 0 \text{ a.s.} \quad (33)$$

To study the asymptotic behavior of the coefficients  $\beta_g$ , we now also introduce some additional projections. Let  $W_l$  denote an element of  $W := (W_1, \dots, W_p)'$  and by  $W_{-l}$  the vector of all elements of  $W$  except the  $l$ -th. We now introduce the projection of  $W_l$  onto  $R_X(X)$  and the remaining regressors  $W_{-l}$ : for any,  $l = 1, \dots, p$ ,

$$W_l = \sum_{g=1}^{n_G} \mathbb{1}\{G = g\} (\tau_{l,g} R_X(X) + W_{-l}' \delta_{l,g}) + v_l, \quad E \left[ v_l \begin{pmatrix} R_X(X) \\ W_{-l} \end{pmatrix} \middle| G \right] = 0 \text{ a.s.}, \quad (34)$$

where  $\tau_{l,g}$  are scalar constants and  $\delta_{l,g}$  are  $(p-1)$ -dimensional vectors of constants.

**Assumption 9.** For  $l = 1, \dots, p$  and  $g = 1, \dots, n_G$ , the random variable  $v_l$  is such that  $E[\mathbb{1}\{G = g\} v_l^2] > 0$ .

Consider the OLS estimator in (17):

$$\begin{pmatrix} \widehat{\rho}_g \\ \widehat{\beta}_g \end{pmatrix} = \left( \sum_{i=1}^n \mathbb{1}\{G_i = c\} \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \mathbb{1}\{G_i = c\} \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} R_i^Y. \quad (35)$$

Let  $\sigma_{v,g}^2$ ,  $h_{0,1,g} := h_{1,g}$ ,  $h_{0,2,g} := h_{2,g}$ , and  $h_{0,3,g} := h_{3,g}$  be as defined in Theorem 2. For  $l = 1, \dots, p$  and  $g = 1, \dots, n_G$ , further define  $\sigma_{v_l,g}^2 := E[\mathbb{1}\{G = g\}v_l^2]$ ,

$$\begin{aligned}\phi_{0,g}(\bar{g}, x, w, y) &:= \frac{1}{\sigma_{v,g}^2} [h_{0,1,g}(\bar{g}, x, w, y) + h_{0,2,g}(x, y) + h_{0,3,g}(x)] \\ \phi_{l,g}(\bar{g}, x, y, w) &:= \frac{1}{\sigma_{v_l,g}^2} [h_{l,1,g}(\bar{g}, x, w, y) + h_{l,2,g}(x, y) + h_{l,3,g}(x)]\end{aligned}$$

and

$$\begin{aligned}h_{l,1,g}(\bar{g}, x, w, y) &:= \mathbb{1}\{\bar{g} = g\}(R_Y(y) - \rho_g R_X(x) - w' \beta_g)(w_l - \tau_{l,g} R_X(x) - w'_{-l} \delta_{l,g}), \\ h_{l,2,g}(x, y) &:= E[\mathbb{1}\{G = g\}(I(y, Y) - \rho_g I(x, X) - W' \beta_g)(W_l - \tau_{l,g} R_X(X) - W'_{-l} \delta_{l,g})], \\ h_{l,3,g}(x) &:= E[\mathbb{1}\{G = g\}(R_Y(Y) - \rho_g R_X(X) - W' \beta_g)(W_l - \tau_{l,g} I(x, X) - W'_{-l} \delta_{l,g})].\end{aligned}$$

Finally, let  $\psi_{i,g} := (\phi_{0,g}(G_i, X_i, W_i, Y_i), \phi_{1,g}(G_i, X_i, W_i, Y_i), \dots, \phi_{p,g}(G_i, X_i, W_i, Y_i))'$  and  $\psi_i := (\phi_{0,1}(G_i, X_i, W_i, Y_i), \dots, \phi_{0,n_G}(G_i, X_i, W_i, Y_i), \phi_{1,1}(G_i, X_i, W_i, Y_i), \dots, \phi_{1,n_G}(G_i, X_i, W_i, Y_i), \dots)'$ .

With the additional assumption and notation, one can derive the following joint asymptotic normality result:

**Theorem 6.** *Suppose (32)–(35) hold and that Assumptions 4–7, and 9 are satisfied. Then:*

$$\sqrt{n} \begin{pmatrix} \widehat{\rho} - \rho \\ \widehat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma),$$

where  $\Sigma := E[\psi_i \psi_i']$  and  $\psi_i$  as defined in this section.

Theorem 2 already shows the asymptotic variance of  $\widehat{\rho}_g$ . Similarly, for the  $l$ -th component of  $\widehat{\beta}_g$ , denoted by  $\widehat{\beta}_{l,g}$ , we have

$$\sqrt{n}(\widehat{\beta}_{l,g} - \beta_{l,g}) \rightarrow_D N(0, \sigma_{\beta_{l,g}}^2),$$

where

$$\sigma_{\beta_{l,g}}^2 := \frac{1}{\sigma_{v_l,g}^4} E \left[ (h_{l,1,g}(G, X, W, Y) + h_{l,2,g}(X, Y) + h_{l,3,g}(X))^2 \right].$$

### D.3 Regression of a General Outcome on a Rank

In this section, we present the asymptotic theory for all coefficients in the rank-rank regression model studied in Section 4.2:

$$Y = \rho R_X(X) + W' \beta + \varepsilon, \quad E \left[ \varepsilon \begin{pmatrix} R_X(X) \\ W \end{pmatrix} \right] = 0, \quad (36)$$

with the projection equations (29) and (30). Consider the OLS estimator in (21):

$$\begin{pmatrix} \widehat{\rho} \\ \widehat{\beta} \end{pmatrix} = \left( \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} \begin{pmatrix} R_i^X & W_i' \end{pmatrix} \right)^{-1} \sum_{i=1}^n \begin{pmatrix} R_i^X \\ W_i \end{pmatrix} Y_i. \quad (37)$$

Let  $\sigma_\nu^2$ ,  $h_{0,1} := h_1$ ,  $h_{0,2} := h_2$ , and  $h_{0,3} := h_3$  be defined as in Theorem 3. For  $l = 1, \dots, p$ , further define  $\sigma_{v_l}^2 := E[v_l^2]$ ,

$$\begin{aligned}\phi_0(x, w, y) &:= \frac{1}{\sigma_\nu^2} [h_{0,1}(x, w, y) + h_{0,2}(x) + h_{0,3}(x)] \\ \phi_l(x, y, w) &:= \frac{1}{\sigma_{v_l}^2} [h_{l,1}(x, w, y) + h_{l,2}(x) + h_{l,3}(x)]\end{aligned}$$

and

$$\begin{aligned}h_{l,1}(x, w, y) &:= (y - \rho R_X(x) - w'\beta)(w_l - \tau_l R_X(x) - w'_{-l}\delta_l), \\ h_{l,2}(x) &:= E[(Y - \rho I(x, X) - W'\beta)(W_l - \tau_l R_X(X) - W'_{-l}\delta_l)], \\ h_{l,3}(x) &:= E[(Y - \rho R_X(X) - W'\beta)(W_l - \tau_l I(x, X) - W'_{-l}\delta_l)].\end{aligned}$$

Finally, let  $\psi_i := (\phi_0(X_i, W_i, Y_i), \phi_1(X_i, W_i, Y_i), \dots, \phi_p(X_i, W_i, Y_i))'$ .

With the additional notation, one can derive the following joint asymptotic normality result:

**Theorem 7.** *Suppose (29), (30), (36), and (37) hold and that Assumptions 1–3, and 8 are satisfied. Then:*

$$\sqrt{n} \begin{pmatrix} \widehat{\rho} - \rho \\ \widehat{\beta} - \beta \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i + o_P(1) \rightarrow_D N(0, \Sigma),$$

where  $\Sigma := E[\psi_i \psi_i']$  and  $\psi_i$  as defined in this section.

Theorem 3 already shows the asymptotic variance of  $\widehat{\rho}$ . Similarly, for the  $l$ -th component of  $\widehat{\beta}_l$  we have

$$\sqrt{n}(\widehat{\beta}_l - \beta_l) \rightarrow_D N(0, \sigma_{\beta_l}^2),$$

where

$$\sigma_{\beta_l}^2 := \frac{1}{\sigma_{v_l}^4} E \left[ (h_{l,1}(X, W, Y) + h_{l,2}(X, Y) + h_{l,3}(X))^2 \right].$$

## E Proofs

### E.1 Proofs for Section 2

Define the sample counterpart to  $M_{kl}$ :

$$\widehat{M}_{kl} := \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j) \right)^k \left( \widehat{F}_Y(Y_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_Y(Y_j) \right)^l.$$

Before proving Lemma 1, we state and prove the following auxiliary lemma.

**Lemma 5.** *Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be an i.i.d. sample from a distribution  $F$  with continuous marginals. Then  $\widehat{M}_{kl} = M_{kl} + o_p(1)$  for  $(k, l) \in \{(0, 2), (2, 0), (2, 2), (3, 1), (4, 0)\}$ .*

*Proof.* Define  $\widehat{P}_{X_i} := \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j)$ ,  $\widetilde{P}_{X_i} := F_X(X_i) - \frac{1}{n} \sum_{j=1}^n F_X(X_j)$ , and  $P_{X_i} := F_X(X_i) - E[F_X(X)]$ . Similarly, define  $\widehat{P}_{Y_i}$ ,  $\widetilde{P}_{Y_i}$ , and  $P_{Y_i}$ . Also, let  $R_n := \max\{\sup_{x \in \mathbb{R}} |\widehat{F}_X(x) - F_X(x)|, \sup_{y \in \mathbb{R}} |\widehat{F}_Y(y) - F_Y(y)|\}$ . Note that by the Glivenko-Cantelli theorem (e.g., Theorem 1.3 in [Dudley \(2014\)](#)),  $R_n = o_p(1)$ . We will use this bound, as well as the elementary identity  $a^2 - b^2 = (a - b)^2 + 2(a - b)b$ , several times without referring to them explicitly.

First, consider

$$\widehat{M}_{20} = \frac{1}{n} \sum_{i=1}^n \widehat{P}_{X_i}^2 = \frac{1}{n} \sum_{i=1}^n \widehat{F}_X(X_i)^2 - \left( \frac{1}{n} \sum_{i=1}^n \widehat{F}_X(X_i) \right)^2.$$

Using the law of large numbers, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \widehat{F}_X(X_i)^2 &= \frac{1}{n} \sum_{i=1}^n F_X(X_i)^2 + \frac{1}{n} \sum_{i=1}^n (\widehat{F}_X(X_i)^2 - F_X(X_i)^2) \\ &= E[F_X(X)^2] + o_p(1) + \frac{1}{n} \sum_{i=1}^n (\widehat{F}_X(X_i) - F_X(X_i))^2 + \frac{2}{n} \sum_{i=1}^n (\widehat{F}_X(X_i) - F_X(X_i))F_X(X_i) \\ &= E[F_X(X)^2] + o_p(1). \end{aligned}$$

Furthermore,

$$\left( \frac{1}{n} \sum_{i=1}^n \widehat{F}_X(X_i) \right)^2 = \left( \frac{1}{n} \sum_{i=1}^n F_X(X_i) + \frac{1}{n} \sum_{i=1}^n (\widehat{F}_X(X_i) - F_X(X_i)) \right)^2 = (E[F_X(X)] + o_p(1))^2.$$

Therefore, we have shown  $\widehat{M}_{20} = M_{20} + o_p(1)$ . Analogously,  $\widehat{M}_{02} = M_{02} + o_p(1)$ .

Second, consider

$$\begin{aligned} \widehat{M}_{22} &= \frac{1}{n} \sum_{i=1}^n \widehat{P}_{X_i}^2 \widehat{P}_{Y_i}^2 = \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^2 - \widetilde{P}_{X_i}^2) \widehat{P}_{Y_i}^2 + \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{Y_i}^2 - \widetilde{P}_{Y_i}^2) \widetilde{P}_{X_i}^2 + \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{X_i}^2 \widetilde{P}_{Y_i}^2 \\ &=: A_n + B_n + C_n. \end{aligned}$$

Here,

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^2 - \widetilde{P}_{X_i}^2) \widehat{P}_{Y_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i} - \widetilde{P}_{X_i})^2 \widehat{P}_{Y_i}^2 + \frac{2}{n} \sum_{i=1}^n (\widehat{P}_{X_i} - \widetilde{P}_{X_i}) \widetilde{P}_{X_i} \widehat{P}_{Y_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_X(X_i) - F_X(X_i) - \frac{1}{n} \sum_{j=1}^n (\widehat{F}_X(X_j) - F_X(X_j)) \right)^2 \widehat{P}_{Y_i}^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^n \left( \widehat{F}_X(X_i) - F_X(X_i) - \frac{1}{n} \sum_{j=1}^n (\widehat{F}_X(X_j) - F_X(X_j)) \right) \widetilde{P}_{X_i} \widehat{P}_{Y_i}^2, \end{aligned}$$

and so

$$|A_n| \leq 4R_n^2 \frac{1}{n} \sum_{i=1}^n \widehat{P}_{Y_i}^2 - 2R_n \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{X_i} \widehat{P}_{Y_i}^2 = o_p(1)$$

since  $|\widehat{P}_{Y_i}| \leq 1$ ,  $|\widetilde{P}_{X_i}| \leq 1$ . Also, by an analogous argument,  $B_n = o_p(1)$ . In addition,

$$\begin{aligned} C_n &= \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{X_i}^2 \widetilde{P}_{Y_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\widetilde{P}_{X_i}^2 - P_{X_i}^2) \widetilde{P}_{Y_i}^2 + \frac{1}{n} \sum_{i=1}^n (\widetilde{P}_{Y_i}^2 - P_{Y_i}^2) P_{X_i}^2 + \frac{1}{n} \sum_{i=1}^n P_{X_i}^2 P_{Y_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\widetilde{P}_{X_i} - P_{X_i})^2 \widetilde{P}_{Y_i}^2 + \frac{2}{n} \sum_{i=1}^n (\widetilde{P}_{X_i} - P_{X_i}) P_{X_i} \widetilde{P}_{Y_i}^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n (\widetilde{P}_{Y_i} - P_{Y_i})^2 P_{X_i}^2 + \frac{2}{n} \sum_{i=1}^n (\widetilde{P}_{Y_i} - P_{Y_i}) P_{Y_i} P_{X_i}^2 + \frac{1}{n} \sum_{i=1}^n P_{X_i}^2 P_{Y_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{Y_i}^2 \left( \frac{1}{n} \sum_{j=1}^n (F_X(X_j) - E(F_X(X))) \right)^2 + \frac{2}{n} \sum_{i=1}^n P_{X_i} \widetilde{P}_{Y_i}^2 \left( \frac{1}{n} \sum_{j=1}^n (F_X(X_j) - E(F_X(X))) \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n P_{X_i}^2 \left( \frac{1}{n} \sum_{j=1}^n (F_Y(Y_j) - E(F_Y(Y))) \right)^2 + \frac{2}{n} \sum_{i=1}^n P_{Y_i} P_{X_i}^2 \left( \frac{1}{n} \sum_{j=1}^n (F_Y(Y_j) - E(F_Y(Y))) \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n (F_X(X_i) - E[F_X(X)])^2 (F_Y(Y_i) - E[F_Y(Y)])^2 \\ &= M_{22} + o_p(1) \end{aligned}$$

by the law of large numbers and the bounds  $|\widetilde{P}_{Y_i}| \leq 1$ ,  $|P_{X_i}^2| \leq 1$ , and  $|P_{Y_i}| \leq 1$ .

Third, consider

$$\begin{aligned} \widehat{M}_{31} &= \frac{1}{n} \sum_{i=1}^n \widehat{P}_{X_i}^3 \widehat{P}_{Y_i} = \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^3 - \widetilde{P}_{X_i}^3) \widehat{P}_{Y_i} + \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{Y_i} - \widetilde{P}_{Y_i}) \widetilde{P}_{X_i}^3 + \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{X_i}^3 \widetilde{P}_{Y_i} \\ &=: D_n + E_n + F_n. \end{aligned}$$

Using the elementary identity  $a^3 - b^3 = (a - b)^3 + 3ab(a - b)$ , we have

$$D_n = \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^3 - \widetilde{P}_{X_i}^3) \widehat{P}_{Y_i} = \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i} - \widetilde{P}_{X_i})^3 \widehat{P}_{Y_i} + \frac{3}{n} \sum_{i=1}^n (\widehat{P}_{X_i} - \widetilde{P}_{X_i}) \widehat{P}_{X_i} \widetilde{P}_{X_i} \widehat{P}_{Y_i}.$$

Then, continue the proof similarly as for  $\widehat{M}_{22}$ .

Finally, consider

$$\widehat{M}_{40} = \frac{1}{n} \sum_{i=1}^n \widehat{P}_{X_i}^4 = \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^2 - \widetilde{P}_{X_i}^2) \widehat{P}_{X_i}^2 + \frac{1}{n} \sum_{i=1}^n (\widehat{P}_{X_i}^2 - \widetilde{P}_{X_i}^2) \widetilde{P}_{X_i}^2 + \frac{1}{n} \sum_{i=1}^n \widetilde{P}_{X_i}^2 \widetilde{P}_{X_i}^2.$$

Then, continue the proof similarly as for  $\widehat{M}_{22}$ .

Q.E.D.

*Proof of Lemma 1.* Consider first the estimator for homoskedastic regression errors  $\sigma_{hom}^2$ . We have

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i^2 &= \frac{1}{n} \sum_{i=1}^n [R_i^Y - \bar{R}^Y - \widehat{\rho}(R_i^X - \bar{R}^X)]^2 \\
&= \frac{1}{n} \sum_{i=1}^n (R_i^Y - \bar{R}^Y)^2 - \widehat{\rho}^2 \frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_Y(Y_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_Y(Y_j) \right)^2 - \widehat{\rho}^2 \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j) \right)^2 \\
&= \widehat{M}_{02} - \widehat{\rho}^2 \widehat{M}_{20} = M_{02} - \rho^2 M_{20} + o_p(1) = M_{02}(1 - \rho^2) + o_p(1)
\end{aligned}$$

where the second-to-last equality follows from Lemma 5 and the last one from noting that  $M_{20} = M_{02}$ . Similarly,

$$\frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2 = M_{02} + o_p(1),$$

and so (7) follows.

Next, consider the Eicker-White estimator  $\sigma_{EW}^2$ . Similarly as above, we have

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i^2 (R_i^X - \bar{R}^X)^2 &= \frac{1}{n} \sum_{i=1}^n [R_i^Y - \bar{R}^Y - \widehat{\rho}(R_i^X - \bar{R}^X)]^2 (R_i^X - \bar{R}^X)^2 \\
&= \frac{1}{n} \sum_{i=1}^n (R_i^Y - \bar{R}^Y)^2 (R_i^X - \bar{R}^X)^2 \\
&\quad - 2\widehat{\rho} \frac{1}{n} \sum_{i=1}^n (R_i^Y - \bar{R}^Y)(R_i^X - \bar{R}^X)^3 \\
&\quad + \widehat{\rho}^2 \frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^4 \\
&= \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_Y(Y_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_Y(Y_j) \right)^2 \left( \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j) \right)^2 \\
&\quad - 2\widehat{\rho} \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_Y(Y_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_Y(Y_j) \right) \left( \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j) \right)^3 \\
&\quad + \widehat{\rho}^2 \frac{1}{n} \sum_{i=1}^n \left( \widehat{F}_X(X_i) - \frac{1}{n} \sum_{j=1}^n \widehat{F}_X(X_j) \right)^4 \\
&= \widehat{M}_{22} - 2\widehat{\rho} \widehat{M}_{31} + \widehat{\rho}^2 \widehat{M}_{40} = M_{22} - 2\rho M_{31} + \rho^2 M_{40} + o_p(1),
\end{aligned}$$

where the last equality follows from Lemma 5. Similarly,

$$\left( \frac{1}{n} \sum_{i=1}^n (R_i^X - \bar{R}^X)^2 \right)^2 = M_{20}^2 + o_p(1),$$

and so (8) follows by noting that  $M_{20} = 1/12$ , and  $M_{40} = 1/80$ .

Q.E.D.

*Proof of Lemma 2.* Let  $a \in (0, 1)$  be a constant and let  $F$  be the cdf of a pair  $(X, Y)$  of  $U([0, 1])$  random variables such that

$$Y = \begin{cases} a - X, & \text{if } 0 \leq X \leq a, \\ X, & \text{if } a < X \leq 1. \end{cases}$$

Then

$$E[\mathbf{1}\{X \leq x\}Y] = \begin{cases} ax - x^2/2, & \text{if } 0 \leq x \leq a, \\ x^2/2, & \text{if } a < x \leq 1. \end{cases}$$

and

$$E[\mathbf{1}\{Y \leq y\}X] = \begin{cases} ay - y^2/2, & \text{if } 0 \leq y \leq a, \\ y^2/2, & \text{if } a < y \leq 1. \end{cases}$$

Hence, given that  $X > a$  if and only if  $Y > a$ , the function  $h$  in (14) satisfies

$$\begin{aligned} E[h(X, Y)] &= E\left[\left(X(a - X) - (aX - X^2/2) - (a(a - X) - (a - X)^2/2)\right)\mathbf{1}\{X \leq a\}\right] \\ &= E[-(a^2/2)\mathbf{1}\{X \leq a\}] = -a^3/2 \end{aligned}$$

and

$$\begin{aligned} E[h(X, Y)^2] &= E\left[\left(X(a - X) - (aX - X^2/2) - (a(a - X) - (a - X)^2/2)\right)^2\mathbf{1}\{X \leq a\}\right] \\ &= E[(a^2/2)^2\mathbf{1}\{X \leq a\}] = a^5/4. \end{aligned}$$

Therefore, given that  $\sigma^2 = 144\text{Var}(h(X, Y))$  by the discussion in a remark after Theorem 1, it follows that

$$\sigma^2 = 144(E[h(X, Y)^2] - (E[h(X, Y)])^2) = 36(a^5 - a^6)$$

On the other hand, by Lemma 1,

$$\begin{aligned} \sigma_{hom}^2 &= 1 - \rho^2 = 1 - (12E[XY] - 3)^2 = -8 - 144(E[XY])^2 + 72E[XY] \\ &= -8 - 144\left(\frac{1}{3} - \frac{a^3}{6}\right)^2 + 72\left(\frac{1}{3} - \frac{a^3}{6}\right) = 4(a^3 - a^6). \end{aligned}$$

In addition, tedious algebra shows that

$$M_{22} = \frac{1}{80} - \frac{a^3}{6} + \frac{a^4}{3} - \frac{a^5}{6}, \quad M_{31} = \frac{1}{80} - \frac{a^3}{8} + \frac{a^4}{4} - \frac{3a^5}{20}, \quad \rho = 1 - 2a^3.$$

Thus, by Lemma 1,

$$\begin{aligned} \sigma_{EW}^2 &= 144\left(\left(\frac{1}{80} - \frac{a^3}{6} + \frac{a^4}{3} - \frac{a^5}{6}\right) \right. \\ &\quad \left. - \left(\frac{1}{40} - \frac{3a^3}{10} + \frac{a^4}{2} - \frac{3a^5}{10} + \frac{a^6}{2} - a^7 + \frac{3a^8}{5}\right) + \left(\frac{1}{80} - \frac{a^3}{20} + \frac{a^6}{20}\right)\right) \\ &= 144\left(\frac{a^3}{12} - \frac{a^4}{6} + \frac{2a^5}{15} - \frac{9a^6}{20} + a^7 - \frac{3a^8}{5}\right). \end{aligned}$$

Combining these bounds, we obtain that  $\sigma_{hom}^2/\sigma^2 \rightarrow \infty$  and  $\sigma_{EW}^2/\sigma^2 \rightarrow \infty$  as  $a \rightarrow 0$ , yielding the asserted claim. Q.E.D.

*Proof of Lemma 3.* Let  $F$  be any distribution with continuous marginals and let  $(X, Y)$  be a pair of random variables with distribution  $F$ . Letting  $F_X$  and  $F_Y$  denote the corresponding marginal distributions, it follows that  $U := F_X(X)$  and  $V := F_Y(Y)$  are both  $U([0, 1])$  random variables. Also,  $F(X, Y) = C(U, V)$ , where  $C$  is the copula of  $F$ . In addition,  $\rho = \text{Corr}(F_X(X), F_Y(Y)) = 12E[UV] - 3$ .

Consider first the case  $\rho \geq 0$ . By Lemma 1,

$$\begin{aligned}\sigma_{hom}^2 &= 1 - \rho^2 = (1 - \rho)(1 + \rho) \geq 1 - \rho = 4 - 12E[UV] \\ &= 4 - E[12UV - 6U^2 - 6V^2 + 6U^2 + 6V^2] \\ &= 4 + 6E[(U - V)^2] - 6E[U^2] - 6E[V^2] = 6E[(U - V)^2].\end{aligned}$$

On the other hand, by (4),

$$\begin{aligned}\sigma^2/9 &= \text{Var} \left( (2U - 1)(2V - 1) + 4 \int_0^1 (C(U, v) - Uv)dv + 4 \int_0^1 (C(u, V) - uV)du \right) \\ &\leq 2E \left[ \left( (2U - 1)(2V - 1) + 4 \int_0^1 (U \wedge v - Uv)dv + 4 \int_0^1 (u \wedge V - uV)du - 1 \right)^2 \right] \quad (38)\end{aligned}$$

$$+ 2E \left[ \left( 4 \int_0^1 (C(U, v) - U \wedge v)dv + 4 \int_0^1 (C(u, V) - u \wedge V)du \right)^2 \right]. \quad (39)$$

Here, observe that

$$\int_0^1 (U \wedge v - Uv)dv = \frac{U - U^2}{2} \quad \text{and} \quad \int_0^1 (u \wedge V - uV)du = \frac{V - V^2}{2}.$$

Thus, the expression in (38) is equal to

$$2E [(4UV - 2U^2 - 2V^2)^2] = 8E [(U - V)^4] \leq 8E [(U - V)^2].$$

To bound (39), we claim that

$$|C(u, v) - u \wedge v| \leq P(|U - V| \geq |u - v|). \quad (40)$$

Indeed, if  $u \leq v$ , then

$$\begin{aligned}C(u, v) - u \wedge v &= E[\mathbf{1}\{U \leq u\}\mathbf{1}\{V \leq v\}] - E[\mathbf{1}\{U \leq u\}\mathbf{1}\{U \leq v\}] \\ &= E[\mathbf{1}\{U \leq u\}(\mathbf{1}\{V \leq v\} - \mathbf{1}\{U \leq v\})] = -E[\mathbf{1}\{U \leq u\}\mathbf{1}\{V > v\}],\end{aligned}$$

and so (40) follows. On the other hand, if  $u > v$ , then

$$\begin{aligned}C(u, v) - u \wedge v &= E[\mathbf{1}\{U \leq u\}\mathbf{1}\{V \leq v\}] - E[\mathbf{1}\{V \leq u\}\mathbf{1}\{V \leq v\}] \\ &= E[\mathbf{1}\{V \leq v\}(\mathbf{1}\{U \leq u\} - \mathbf{1}\{V \leq u\})] = -E[\mathbf{1}\{V \leq v\}\mathbf{1}\{U > u\}],\end{aligned}$$



and so (40) follows as well, yielding the claim. Thus,

$$\begin{aligned} \int_0^1 |C(u, v) - u \wedge v| dv &\leq \int_0^1 P(|U - V| \geq |u - v|) dv \\ &\leq 2 \int_0^1 P(|U - V| \geq t) dt \leq 2E[|U - V|] \end{aligned}$$

and, similarly,

$$\int_0^1 |C(u, v) - u \wedge v| du \leq 2E[|U - V|].$$

Hence, the expression in (39) is bounded from above by

$$2 \times 16^2 (E[|U - V|])^2 \leq 512E[|U - V|^2].$$

Therefore,  $\sigma^2 \leq 9 \times 520E[|U - V|^2]$ , and so  $\sigma_{hom}^2/\sigma^2$  is bounded below from zero.

In turn, the case  $\rho < 0$  can be treated similarly. The asserted claim follows. Q.E.D.

## E.2 Proofs for Section 3

The proof of Theorem 1 relies on a few auxiliary results, which are presented below. As in the main text, throughout this appendix,  $\omega \in [0, 1]$  is fixed and the same in the definitions of  $R_X$ ,  $\widehat{R}_X$ ,  $R_Y$ , and  $\widehat{R}_Y$ .

*Proof of Theorem 1.* By the Frisch-Waugh-Lovell theorem, the estimator  $\widehat{\rho}$  in (11) can alternatively be written as

$$\widehat{\rho} = \frac{\sum_{i=1}^n R_i^Y (R_i^X - W_i' \widehat{\gamma})}{\sum_{i=1}^n (R_i^X - W_i' \widehat{\gamma})^2} = \frac{\sum_{i=1}^n \widehat{R}_Y(Y_i) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})}{\sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2}, \quad (41)$$

where

$$\widehat{\gamma} = \left( \sum_{i=1}^n W_i W_i' \right)^{-1} \left( \sum_{i=1}^n W_i \widehat{R}_X(X_i) \right) \quad (42)$$

is the OLS estimator of a regression of  $\widehat{R}_X(X_i)$  on  $W_i$ . Therefore, using  $\sum_{i=1}^n W_i (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}) = 0$  and replacing  $\widehat{R}_Y(Y_i)$  in the numerator of (41) by  $\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) + \rho \widehat{R}_X(X_i)$ ,

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})}{\frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2}. \quad (43)$$

Thus, by Assumption 3 and Lemmas 9 and 10,

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})}{\frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2} + o_P(1). \quad (44)$$

Consider the numerator. Define  $Z_i := (Y_i, X_i, W_i)'$  for all  $i = 1, \dots, n$  and

$$f(Z_i, Z_j, Z_k) := (I(Y_j, Y_i) - \rho I(X_j, X_i) - W_i' \beta) (I(X_k, X_i) - W_i' \widehat{\gamma})$$

for all  $i, j, k = 1, \dots, n$ . Also, define  $h(Z_i, Z_j, Z_k) := 6^{-1} \sum_{i_1, i_2, i_3} f(Z_{i_1}, Z_{i_2}, Z_{i_3})$ , where the sum is over all six permutations of the triplet  $(i, j, k)$ , for all  $i, j, k = 1, \dots, n$ . Note that  $h$  is a symmetric function satisfying  $E[h(Z_i, Z_j, Z_k)] = 0$  whenever  $1 \leq i < j < k \leq n$ . Moreover, let  $\Sigma_1$  be the set of triplets  $(i, j, k)$  in  $\{1, \dots, n\}^3$  such that all three elements are the same,  $\Sigma_2$  be the set of triplet  $(i, j, k)$  in  $\{1, \dots, n\}^3$  such that two out of three elements are the same, and  $\Sigma_3$  be the set of triplet  $(i, j, k)$  in  $\{1, \dots, n\}^3$  such that all three elements are different. It is then easy to check that for all  $l = 1, 2, 3$ ,

$$\sum_{(i,j,k) \in \Sigma_l} f(Z_i, Z_j, Z_k) = \sum_{(i,j,k) \in \Sigma_l} h(Z_i, Z_j, Z_k),$$

and so

$$\sum_{i,j,k=1}^n f(Z_i, Z_j, Z_k) = \sum_{i,j,k=1}^n h(Z_i, Z_j, Z_k).$$

Therefore,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) (\widehat{R}_X(X_i) - W_i' \gamma) \\ &= \frac{1}{n^3} \sum_{i,j,k=1}^n (I(Y_j, Y_i) - \rho I(X_j, X_i) - W_i' \beta) (I(X_k, X_i) - W_i' \gamma) + o_p(n^{-1/2}) \\ &= \frac{1}{n^3} \sum_{i,j,k=1}^n f(Z_i, Z_j, Z_k) + o_p(n^{-1/2}) = \frac{1}{n^3} \sum_{i,j,k=1}^n h(Z_i, Z_j, Z_k) + o_p(n^{-1/2}) \\ &= \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} h(Z_i, Z_j, Z_k) + o_p(n^{-1/2}), \end{aligned}$$

where the last line follows from the Lemma on p. 206 of [Serfling \(2002\)](#), whose application is justified since  $E(h(Z_i, Z_j, Z_k)^2) < \infty$  by Assumption 2. Furthermore, the results on p. 188 of [Serfling \(2002\)](#) imply that the U-statistic can be projected onto the basic observations. To compute the projection, note that, for  $z = (y, x, w)'$ ,

$$\begin{aligned} E[f(Z_1, Z_2, Z_3) | Z_1 = z] &= E[f(Z_1, Z_3, Z_2) | Z_1 = z] \\ &= (R_Y(y) - \rho R_X(x) - w' \beta)(R_X(x) - w' \gamma) = h_1(x, w, y), \\ E[f(Z_2, Z_1, Z_3) | Z_1 = z] &= E[f(Z_3, Z_1, Z_2) | Z_1 = z] \\ &= E[(I(y, Y) - \rho I(x, X) - W' \beta)(R_X(X) - W' \gamma)] = h_2(x, y), \\ E[f(Z_2, Z_3, Z_1) | Z_1 = z] &= E[f(Z_3, Z_2, Z_1) | Z_1 = z] \\ &= E[(R_Y(Y) - \rho R_X(X) - W' \beta)(I(x, X) - W' \gamma)] = h_3(x). \end{aligned}$$

Denoting  $\tilde{h}(z) = E[h(Z_1, Z_2, Z_3) | Z_1 = z]$ , we thus have  $\tilde{h}(z) = (h_1(x, w, y) + h_2(x, y) + h_3(x))/3$ ,

and so

$$\begin{aligned} \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} h(Z_i, Z_j, Z_k) &= \frac{3}{n} \sum_{i=1}^n \tilde{h}(Z_i) + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ h_1(X_i, W_i, Y_i) + h_2(X_i, Y_i) + h_3(X_i) \right\} + o_P(n^{-1/2}). \end{aligned}$$

Therefore, the central limit theorem (e.g., Theorem 9.5.6 in [Dudley \(2002\)](#)) implies that the numerator of (44) is asymptotically normal with mean zero and variance  $E[(h_1(X, W, Y) + h_2(X, Y) + h_3(X))^2]$ . By Lemma 9, the denominator of (44) converges in probability to  $\sigma_\nu^2$ , so that Slutsky's lemma yields the asserted claim. Q.E.D.

*Proof of Lemma 4.* First, we prove that

$$\left| \frac{1}{n} \sum_{i=1}^n (H_{1i} + H_{2i} + H_{3i})^2 - \frac{1}{n} \sum_{i=1}^n (h_1(X_i, W_i, Y_i) + h_2(X_i, Y_i) + h_3(X_i))^2 \right| = o_P(1). \quad (45)$$

To do so, observe that  $E[(h_1(X, W, Y) + h_2(X, Y) + h_3(X))^2] < \infty$  by Assumption 2. Hence, it follows from Lemma 11 that (45) holds if

$$R_n := \frac{1}{n} \sum_{i=1}^n (H_{1i} + H_{2i} + H_{3i} - h_1(X_i, W_i, Y_i) - h_2(X_i, Y_i) - h_3(X_i))^2 = o_P(1). \quad (46)$$

In turn, by the triangle inequality,  $R_n \leq 9(R_{1n} + R_{2n} + R_{3n})$ , where

$$\begin{aligned} R_{1n} &:= \frac{1}{n} \sum_{i=1}^n (H_{1i} - h_1(X_i, W_i, Y_i))^2, \\ R_{2n} &:= \frac{1}{n} \sum_{i=1}^n (H_{2i} - h_2(X_i, Y_i))^2, \\ R_{3n} &:= \frac{1}{n} \sum_{i=1}^n (H_{3i} - h_3(X_i))^2. \end{aligned}$$

We bound these three terms in turn.

Regarding  $R_{1n}$ , we have  $E[|R_Y(Y) - \rho R_X(X) - W'\beta|^4] < \infty$  and  $E[|R_X(X) - W'\gamma|^4] < \infty$  by Assumption 2. Also,

$$\frac{1}{n} \sum_{i=1}^n \left| (R_i^Y - \hat{\rho} R_i^X - W_i' \hat{\beta}) - (R_Y(Y_i) - \rho R_X(X_i) - W_i' \beta) \right|^4 = o_P(1)$$

by Lemma 7, Theorems 1 and 5 and Assumption 2. In addition,

$$\frac{1}{n} \sum_{i=1}^n \left| (R_i^X - W_i' \hat{\gamma}) - (R_X(X_i) - W_i' \gamma) \right|^4 = o_P(1)$$

by Lemmas 6 and 7 and Assumption 2. Hence,  $R_{1n} = o_P(1)$  by Lemma 12.

Regarding  $R_{2n}$ , we have by the triangle inequality that

$$\begin{aligned}\sqrt{R_{2n}} &\leq \sup_{x,y \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (I(y, Y_j) - \widehat{\rho}I(x, X_j) - W'_j \widehat{\beta})(R_j^X - W'_j \widehat{\gamma}) - h_2(x, y) \right| \\ &\leq \sup_{x,y \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (I(y, Y_j) - \widehat{\rho}I(x, X_j) - W'_j \widehat{\beta})(R_j^X - W'_j \widehat{\gamma}) - \widehat{h}_2(x, y) \right| \\ &\quad + \sup_{x,y \in \mathbb{R}} \left| \widehat{h}_2(x, y) - h_2(x, y) \right|,\end{aligned}$$

where

$$\widehat{h}_2(x, y) := \frac{1}{n} \sum_{j=1}^n (I(y, Y_j) - \rho I(x, X_j) - W'_j \beta)(R_X(X_j) - W'_j \gamma), \quad \text{for all } x, y \in \mathbb{R}.$$

Also,  $\sup_{x,y \in \mathbb{R}} |\widehat{h}_2(x, y) - h_2(x, y)| = o_P(1)$  under Assumption 2 by Lemma 8 and the triangle inequality. In addition, for all  $j = 1, \dots, n$ , denote  $B_j(x, y) := I(y, Y_j) - \rho I(x, X_j) - W'_j \beta$ ,  $\widehat{B}_j(x, y) := I(y, Y_j) - \widehat{\rho}I(x, X_j) - W'_j \widehat{\beta}$ , for all  $x, y \in \mathbb{R}$ ,  $C_j := R_X(X_j) - W'_j \gamma$ , and  $\widehat{C}_j := R_j^X - W'_j \widehat{\gamma}$ . Then  $\sup_{x,y \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (\widehat{B}_j(x, y) - B_j(x, y))^4 = o_P(1)$  by Theorems 1 and 5 and Assumption 2;  $\frac{1}{n} \sum_{j=1}^n (\widehat{C}_j - C_j)^4 = o_P(1)$  by Lemmas 6 and 7 and Assumption 2;  $\sup_{x,y \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n B_j(x, y)^4 = O_P(1)$  and  $\frac{1}{n} \sum_{j=1}^n C_j^4 = O_P(1)$  by Assumption 2. Hence, by Lemma 12,

$$\begin{aligned}&\sup_{x,y \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (I(y, Y_j) - \widehat{\rho}I(x, X_j) - W'_j \widehat{\beta})(R_j^X - W'_j \widehat{\gamma}) - \widehat{h}_2(x, y) \right| \\ &= \sup_{x,y \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n \widehat{B}_j(x, y) \widehat{C}_j - B_j(x, y) C_j \right| \leq \sup_{x,y \in \mathbb{R}} \sqrt{\frac{1}{n} \sum_{j=1}^n (\widehat{B}_j(x, y) \widehat{C}_j - B_j(x, y) C_j)^2} = o_P(1).\end{aligned}$$

Combining the presented bounds gives  $R_{2n} = o_P(1)$ .

Regarding  $R_{3n}$ , we have by the triangle inequality that

$$\begin{aligned}\sqrt{R_{3n}} &\leq \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (R_j^Y - \widehat{\rho}R_j^X - W'_j \widehat{\beta})(I(x, X_j) - W'_j \widehat{\gamma}) - h_3(x) \right| \\ &\leq \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (R_j^Y - \widehat{\rho}R_j^X - W'_j \widehat{\beta})(I(x, X_j) - W'_j \widehat{\gamma}) - \widehat{h}_3(x) \right| + \sup_{x \in \mathbb{R}} |\widehat{h}_3(x) - h_3(x)|,\end{aligned}$$

where

$$\widehat{h}_3(x) := \frac{1}{n} \sum_{j=1}^n (R_Y(Y_j) - \rho R_X(X_j) - W'_j \beta)(I(x, X_j) - W'_j \gamma), \quad \text{for all } x \in \mathbb{R}.$$

Also,  $\sup_{x \in \mathbb{R}} |\widehat{h}_3(x) - h_3(x)| = o_P(1)$  under Assumption 2 by Lemma 8 and the triangle inequality. In addition, for all  $j = 1, \dots, n$ , denote  $B_j := R_Y(Y_j) - \rho R_X(X_j) - W'_j \beta$ ,  $\widehat{B}_j := R_j^Y - \widehat{\rho}R_j^X - W'_j \widehat{\beta}$ ,

$C_j(x) := I(x, X_j) - W_j' \gamma$ ,  $\widehat{C}_j(x) := I(x, X_j) - W_j' \widehat{\gamma}$ , for all  $x \in \mathbb{R}$ . Then  $\frac{1}{n} \sum_{j=1}^n (\widehat{B}_j - B_j)^4 = o_P(1)$  by Theorems 1 and 5, Lemma 7, and Assumption 2;  $\sup_{x \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (\widehat{C}_j(x) - C_j(x))^4 = o_P(1)$  by Lemma 6 and Assumption 2;  $\frac{1}{n} \sum_{j=1}^n B_j^4 = O_P(1)$  and  $\sup_{x \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n C_j(x)^4 = O_P(1)$  by Assumption 2. Hence, by Lemma 12,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (R_j^Y - \widehat{\rho} R_j^X - W_j' \widehat{\beta})(I(x, X_j) - W_j' \widehat{\gamma}) - \widehat{h}_3(x) \right| \\ &= \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n \widehat{B}_j \widehat{C}_j(x) - B_j C_j(x) \right| \leq \sup_{x \in \mathbb{R}} \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{B}_i \widehat{C}_i(x) - B_i C_i(x))^2} = o_P(1). \end{aligned}$$

Combining presented bounds gives  $R_{3n} = o_P(1)$ .

Thus, (46) and hence (45) are satisfied. In turn, combining (45) with the law of large numbers yields

$$\frac{1}{n} \sum_{i=1}^n (H_{1i} + H_{2i} + H_{3i})^2 \rightarrow_P E[(h_1(X, W, Y) + h_2(X, Y) + h_3(X))^2]. \quad (47)$$

Also,  $\widehat{\sigma}_\nu^2 \rightarrow_P \sigma_\nu^2$  by Lemma 9. The asserted claim now follows from combining these convergence results with the continuous mapping theorem. Q.E.D.

**Lemma 6.** *Under Assumptions 1 and 2, we have  $\sqrt{n}(\widehat{\gamma} - \gamma) = O_P(1)$ .*

*Proof.* By (12) and (42),

$$\sqrt{n}(\widehat{\gamma} - \gamma) = \left( \frac{1}{n} \sum_{i=1}^n W_i W_i' \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i (\nu_i + \widehat{R}_X(X_i) - R_X(X_i)) \right).$$

By the law of large numbers and Assumption 2,  $n^{-1} \sum_{i=1}^n W_i W_i' \rightarrow_P E[WW']$ . Also, by (12), Assumption 2, and Chebyshev's inequality,  $n^{-1/2} \sum_{i=1}^n W_i \nu_i = O_P(1)$ . In addition,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i (\widehat{R}_X(X_i) - R_X(X_i)) &= \frac{1}{n^{3/2}} \sum_{i,j=1}^n W_i (I(X_j, X_i) - R_X(X_i)) + o_p(n^{-1/2}) \\ &= \frac{1}{n^{3/2}} \sum_{i=1}^n W_i (\omega - R_X(X_i)) + \frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j \neq i} W_i (I(X_j, X_i) - R_X(X_i)) + o_p(n^{-1/2}), \end{aligned}$$

where  $E[\|\sum_{i=1}^n W_i (\omega - R_X(X_i))\|] \leq \sum_{i=1}^n E[\|W_i\|] = O(n)$  by Assumption 2 and

$$E \left[ \left\| \sum_{i=1}^n \sum_{j \neq i} W_i (I(X_j, X_i) - R_X(X_i)) \right\|^2 \right] = O(n^3)$$

by Assumption 2 and results on p. 183 in Serfling (2002). Combining the presented bounds gives the asserted claim. Q.E.D.

**Lemma 7.** *Under Assumption 1, we have  $\sup_{x \in \mathbb{R}} |\widehat{R}_X(x) - R_X(x)| = o_P(1)$  and  $\sup_{y \in \mathbb{R}} |\widehat{R}_Y(y) - R_Y(y)| = o_P(1)$ .*

*Proof.* By the Glivenko-Cantelli theorem (e.g., Theorem 1.3 in [Dudley \(2014\)](#)),  $\sup_{x \in \mathbb{R}} |\widehat{F}_X(x) - F_X(x)| = o_P(1)$ . Also, by the Glivenko-Cantelli theorem applied to  $-X_i$ 's instead of  $X_i$ 's,  $\sup_{x \in \mathbb{R}} |\widehat{F}_X^-(x) - F_X^-(x)| = o_P(1)$ . The first claim thus follows. The second claim follows from the same argument. Q.E.D.

**Lemma 8.** *Under Assumptions 1 and 2, we have*

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(y, Y_i)(R_X(X_i) - W_i' \gamma) - E[I(y, Y)(R_X(X) - W' \gamma)] \right| = o_P(1), \quad (48)$$

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(x, X_i)(R_X(X_i) - W_i' \gamma) - E[I(x, X)(R_X(X) - W' \gamma)] \right| = o_P(1), \quad (49)$$

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n (R_Y(Y_i) - \rho R_X(X_i) - W_i' \beta) I(x, X_i) \right. \\ \left. - E[(R_Y(Y) - \rho R_X(X) - W' \beta) I(x, X)] \right| = o_P(1). \end{aligned} \quad (50)$$

*Proof.* For any random variable  $A$ , let  $A^+$  and  $A^-$  denote  $A \mathbf{1}\{A \geq 0\}$  and  $-A \mathbf{1}\{A < 0\}$ , respectively. Then

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(y, Y_i)(R_X(X_i) - W_i' \gamma)^+ - E[I(y, Y)(R_X(X) - W' \gamma)^+] \right| = o_P(1)$$

and

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(y, Y_i)(R_X(X_i) - W_i' \gamma)^- - E[I(y, Y)(R_X(X) - W' \gamma)^-] \right| = o_P(1)$$

by the argument parallel to that used in the proof of the Glivenko-Cantelli theorem (e.g., Theorem 1.3 in [Dudley \(2014\)](#)). Combining these bounds gives (48). In turn, (49) and (50) follow from the same argument. Q.E.D.

**Lemma 9.** *Under Assumptions 1 and 2, we have*

$$\frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2 \rightarrow_P E[(R_X(X) - W' \gamma)^2] = \sigma_v^2.$$

*Proof.* By Assumption 2 and the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n (R_X(X_i) - W_i' \gamma)^2 \rightarrow_P E[(R_X(X) - W' \gamma)^2].$$

Also, using an elementary identity  $a^2 - b^2 = (a - b)^2 + 2(a - b)b$ , we have

$$\begin{aligned} & \left| n^{-1} \sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2 - n^{-1} \sum_{i=1}^n (R_X(X_i) - W_i' \gamma)^2 \right| \\ & \leq \frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - R_X(X_i) - W_i' \widehat{\gamma} + W_i' \gamma)^2 \end{aligned} \quad (51)$$

$$+ \left| \frac{2}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - R_X(X_i) - W_i' \widehat{\gamma} + W_i' \gamma)(R_X(X_i) - W_i' \gamma) \right|. \quad (52)$$

The term in (51) is bounded from above by

$$\begin{aligned} & \frac{2}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - R_X(X_i))^2 + \frac{2}{n} \sum_{i=1}^n (W_i' \widehat{\gamma} - W_i' \gamma)^2 \\ & \leq 2 \sup_{x \in \mathbb{R}} |\widehat{R}_X(x) - R_X(x)|^2 + 2 \|\widehat{\gamma} - \gamma\|^2 \times \left\| \frac{1}{n} \sum_{i=1}^n W_i W_i' \right\| = o_P(1) \end{aligned}$$

by Lemmas 6 and 7 and Assumption 2. The term in (52) is bounded from above by

$$2 \left( \frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - R_X(X_i) - W_i' \widehat{\gamma} + W_i' \gamma)^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n (R_X(X_i) - W_i' \gamma)^2 \right)^{1/2} = o_P(1)$$

by the arguments above. The asserted claim thus follows. Q.E.D.

**Lemma 10.** *Under Assumptions 1 and 2, we have*

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}) \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) (\widehat{R}_X(X_i) - W_i' \gamma) + o_P(1). \end{aligned}$$

*Proof.* Using  $\sum_{i=1}^n W_i (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}) = 0$ , we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}) \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}) \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) (\widehat{R}_X(X_i) - W_i' \gamma) - R_n \end{aligned}$$

where

$$R_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \rho \widehat{R}_X(X_i) - W_i' \beta) W_i' (\widehat{\gamma} - \gamma) = R_{1n} + R_{2n} + R_{3n}$$

with

$$\begin{aligned}
R_{1n} &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - R_Y(Y_i)) W_i' (\widehat{\gamma} - \gamma) \\
R_{2n} &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\rho R_X(X_i) - \rho \widehat{R}_X(X_i)) W_i' (\widehat{\gamma} - \gamma) \\
R_{3n} &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i W_i' (\widehat{\gamma} - \gamma)
\end{aligned}$$

Consider each of these remainder terms separately. First,

$$|R_{1n}| \leq \sup_{y \in \mathbb{R}} |\widehat{R}_Y(y) - R_Y(y)| \times \frac{1}{n} \sum_{i=1}^n \|W_i\| \times \|\sqrt{n}(\widehat{\gamma} - \gamma)\| = o_P(1)$$

by Lemmas 6 and 7 and Assumption 2. Similarly,  $|R_{2n}| = o_P(1)$ . Finally,

$$|R_{3n}| \leq \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i W_i \right\| \times \|\sqrt{n}(\widehat{\gamma} - \gamma)\| = o_P(1)$$

by Lemma 6, Assumption 2, and Chebyshev's inequality. The asserted claim thus follows. Q.E.D.

**Lemma 11.** *For any vectors  $A = (A_1, \dots, A_n)'$  and  $\widehat{A} = (\widehat{A}_1, \dots, \widehat{A}_n)'$ , we have*

$$\left| \frac{1}{n} \sum_{i=1}^n (\widehat{A}_i^2 - A_i^2) \right| \leq \frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i)^2 + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n A_i^2}.$$

*Proof.* We have

$$\left| \frac{1}{n} \sum_{i=1}^n (\widehat{A}_i^2 - A_i^2) \right| = \left| \frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i)(\widehat{A}_i + A_i) \right| \leq \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i + A_i)^2}$$

by the Cauchy-Schwarz inequality. Also,

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i + A_i)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i + 2A_i)^2} \leq \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{A}_i - A_i)^2} + 2 \sqrt{\frac{1}{n} \sum_{i=1}^n A_i^2}$$

by the triangle inequality. Combining these bounds gives the asserted claim. Q.E.D.

**Lemma 12.** *For any vectors  $B = (B_1, \dots, B_n)'$ ,  $\widehat{B} = (\widehat{B}_1, \dots, \widehat{B}_n)'$ ,  $C = (C_1, \dots, C_n)'$ , and  $\widehat{C} = (\widehat{C}_1, \dots, \widehat{C}_n)'$ , we have*

$$\begin{aligned}
\sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{B}_i \widehat{C}_i - B_i C_i)^2} &\leq \left( \frac{1}{n} \sum_{i=1}^n (\widehat{B}_i - B_i)^4 \right)^{1/4} \left( \frac{1}{n} \sum_{i=1}^n (\widehat{C}_i - C_i)^4 \right)^{1/4} \\
&\quad + \left( \frac{1}{n} \sum_{i=1}^n C_i^4 \right)^{1/4} \left( \frac{1}{n} \sum_{i=1}^n (\widehat{B}_i - B_i)^4 \right)^{1/4} \\
&\quad + \left( \frac{1}{n} \sum_{i=1}^n B_i^4 \right)^{1/4} \left( \frac{1}{n} \sum_{i=1}^n (\widehat{C}_i - C_i)^4 \right)^{1/4}.
\end{aligned}$$



*Proof.* For all  $i = 1, \dots, n$ , we have

$$\widehat{B}_i \widehat{C}_i - B_i C_i = (\widehat{B}_i - B_i) \widehat{C}_i + B_i (\widehat{C}_i - C_i) = (\widehat{B}_i - B_i) (\widehat{C}_i - C_i) + C_i (\widehat{B}_i - B_i) + B_i (\widehat{C}_i - C_i).$$

Hence,

$$\begin{aligned} \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{B}_i \widehat{C}_i - B_i C_i)^2} &\leq \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{B}_i - B_i)^2 (\widehat{C}_i - C_i)^2} \\ &\quad + \sqrt{\frac{1}{n} \sum_{i=1}^n C_i^2 (\widehat{B}_i - B_i)^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n B_i^2 (\widehat{C}_i - C_i)^2} \end{aligned}$$

by the triangle inequality. Applying the Cauchy-Schwarz inequality to each term on the right-hand side of this bound yields the asserted claim. Q.E.D.

### E.3 Proofs for Section 4

#### E.3.1 Proofs for Section 4.1

*Proof of Theorem 2.* Fix  $g \in \{1, \dots, n_G\}$  and let  $\widetilde{R}_{i,g}^X := \mathbf{1}\{G_i = g\} R_i^X$ ,  $\widetilde{R}_{i,g}^Y := \mathbf{1}\{G_i = g\} R_i^Y$ , and  $\widetilde{W}_{i,g} := \mathbf{1}\{G_i = g\} W_i$ . Then,  $\widehat{\rho}_g$  is the OLS estimator from a regression of  $\widetilde{R}_{i,g}^Y$  on  $\widetilde{R}_{i,g}^X$  and  $\widetilde{W}_{i,g}$  using all observations  $i = 1, \dots, n$ . Therefore, as in the proof of Theorem 1,

$$\begin{aligned} \sqrt{n}(\widehat{\rho}_g - \rho_g) &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widetilde{R}_{i,g}^Y - \rho_g \widetilde{R}_{i,g}^X) (\widetilde{R}_{i,g}^X - \widetilde{W}_{i,g}' \widehat{\gamma}_g)}{\frac{1}{n} \sum_{i=1}^n (\widetilde{R}_{i,g}^X - \widetilde{W}_{i,g}' \widehat{\gamma}_g)^2} \\ &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}\{G_i = g\} (\widehat{R}_Y(Y_i) - \rho_g \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}_g)}{\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{G_i = g\} (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}_g)^2}, \end{aligned}$$

see (43). By Lemmas 13 and 14, we have

$$\sqrt{n}(\widehat{\rho}_g - \rho_g) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}\{G_i = g\} (\widehat{R}_Y(Y_i) - \rho_g \widehat{R}_X(X_i) - W_i' \beta_g) (\widehat{R}_X(X_i) - W_i' \gamma_g)}{\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{G_i = g\} (\widehat{R}_X(X_i) - W_i' \gamma_g)^2} + o_P(1). \quad (53)$$

Define  $Z_i := (G_i, Y_i, X_i, W_i')'$  for all  $i = 1, \dots, n$  and

$$f_g(Z_i, Z_j, Z_k) := \mathbf{1}\{G_i = g\} (I(Y_j, Y_i) - \rho I(X_j, X_i) - W_i' \beta) (I(X_k, X_i) - W_i' \gamma)$$

for all  $i, j, k = 1, \dots, n$ . Also, define  $h_g(Z_i, Z_j, Z_k) := 6^{-1} \sum_{i_1, i_2, i_3} f_g(Z_{i_1}, Z_{i_2}, Z_{i_3})$ , where the sum is over all six permutations  $(i_1, i_2, i_3)$  of the triplet  $(i, j, k)$ . Note that  $h_g$  is a symmetric function satisfying  $E[h_g(Z_i, Z_j, Z_k)] = 0$  whenever  $1 \leq i < j < k \leq n$ . Also, Assumption 5

implies that  $E(h_g(Z_i, Z_j, Z_k)^2) < \infty$ . Then, letting  $z := (\bar{g}, y, x, w)'$ ,

$$\begin{aligned}
E[f_g(Z_1, Z_2, Z_3) \mid Z_1 = z] &= E[f_g(Z_1, Z_3, Z_2) \mid Z_1 = z] \\
&= \mathbb{1}\{\bar{g} = g\}(R_Y(y) - \rho_g R_X(x) - w' \beta_g)(R_X(x) - w' \gamma_g) \\
&= h_{1,g}(\bar{g}, x, w, y), \\
E[f_g(Z_2, Z_1, Z_3) \mid Z_1 = z] &= E[f_g(Z_3, Z_1, Z_2) \mid Z_1 = z] \\
&= E[\mathbb{1}\{G = g\}(I(y, Y) - \rho_g I(x, X) - W' \beta_g)(R_X(X) - W' \gamma_g)] \\
&= h_{2,g}(x, y), \\
E[f_g(Z_2, Z_3, Z_1) \mid Z_i = z] &= E[f_g(Z_3, Z_2, Z_1) \mid Z_1 = z] \\
&= E[\mathbb{1}\{G = g\}(R_Y(Y) - \rho_g R_X(X) - W' \beta_g)(I(x, X) - W' \gamma_g)] \\
&= h_{3,g}(x),
\end{aligned}$$

we can argue as in the proof of Theorem 1 that

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{G_i = g\} \left( \widehat{R}_Y(Y_i) - \rho_g \widehat{R}_X(X_i) - W_i' \beta_g \right) \left( \widehat{R}_X(X_i) - W_i' \gamma_g \right) \\
= \frac{1}{n} \sum_{i=1}^n \left\{ h_{1,g}(X_i, W_i, Y_i) + h_{2,g}(X_i, Y_i) + h_{3,g}(X_i) \right\} + o_P(n^{-1/2}).
\end{aligned}$$

The remainder of the proof then follows that of Theorem 1. Q.E.D.

**Lemma 13.** *Under Assumption 5, we have for any  $g = 1, \dots, n_G$ ,*

$$n^{-1} \sum_{i=1}^n \mathbb{1}\{G_i = g\} (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}_g)^2 \rightarrow_P E[\mathbb{1}\{G = g\} (R_X(X) - W' \gamma_g)^2] = \sigma_{\nu,g}^2.$$

*Proof.* Similar to that of Lemma 9. Q.E.D.

**Lemma 14.** *Under Assumption 5, we have for any  $g = 1, \dots, n_G$ ,*

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}\{G_i = g\} (\widehat{R}_Y(Y_i) - \rho_g \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma}_g) \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}\{G_i = g\} (\widehat{R}_Y(Y_i) - \rho_g \widehat{R}_X(X_i) - W_i' \beta_g) (\widehat{R}_X(X_i) - W_i' \gamma_g) + o_P(1).
\end{aligned}$$

*Proof.* Similar to that of Lemma 10. Q.E.D.

### E.3.2 Proofs for Section 4.2

*Proof of Theorem 3.* As in the proof of Theorem 1,

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \widehat{R}_X(X_i)) (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})}{\frac{1}{n} \sum_{i=1}^n (\widehat{R}_X(X_i) - W_i' \widehat{\gamma})^2}.$$

By Lemmas 9 and 15, we have

$$\sqrt{n}(\hat{\rho} - \rho) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i) - W_i' \beta)(\hat{R}_X(X_i) - W_i' \gamma)}{\frac{1}{n} \sum_{i=1}^n (\hat{R}_X(X_i) - W_i' \hat{\gamma})^2} + o_P(1). \quad (54)$$

Define  $Z_i := (Y_i, X_i, W_i)'$  for all  $i = 1, \dots, n$  and

$$f(Z_i, Z_j, Z_k) := (Y_i - \rho I(X_j, X_i) - W_i' \beta)(I(X_k, X_i) - W_i' \gamma)$$

for all  $i, j, k = 1, \dots, n$ . Also, define  $h(Z_i, Z_j, Z_k) := 6^{-1} \sum_{i_1, i_2, i_3} f(Z_{i_1}, Z_{i_2}, Z_{i_3})$ , where the sum is taken over all six permutations  $(i_1, i_2, i_3)$  of the triplet  $(i, j, k)$ . Since  $E[\varepsilon^4] < \infty$  and Assumption 2 imply  $E[h(Z_i, Z_j, Z_k)^2] < \infty$ , we can argue as in the proof of Theorem 1 to show that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i) - W_i' \beta)(\hat{R}_X(X_i) - W_i' \gamma) \\ = \frac{1}{n} \sum_{i=1}^n \left\{ h_1(X_i, W_i, Y_i) + h_2(X_i) + h_3(X_i) \right\} + o_P(n^{-1/2}) \end{aligned} \quad (55)$$

where for  $z = (x, y, w)'$ , we define

$$\begin{aligned} E[f(Z_i, Z_j, Z_k) \mid Z_i = z] &= E[f(Z_i, Z_k, Z_j) \mid Z_i = z] \\ &= (y - \rho R_X(x) - w' \beta)(R_X(x) - w' \gamma) = h_1(x, w, y), \\ E[f(Z_j, Z_i, Z_k) \mid Z_i = z] &= E[f(Z_k, Z_i, Z_j) \mid Z_i = z] \\ &= E[(Y - \rho I(x, X) - W' \beta)(R_X(X) - W' \gamma)] = h_2(x), \\ E[f(Z_j, Z_k, Z_i) \mid Z_i = z] &= E[f(Z_k, Z_j, Z_i) \mid Z_i = z] \\ &= E[(Y - \rho R_X(X) - W' \beta)(I(x, X) - W' \gamma)] = h_3(x) \end{aligned}$$

Combining (54) with (55) then implies the desired result by the same argument as that in the proof of Theorem 1. Q.E.D.

**Lemma 15.** *Under Assumption 2, we have*

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i))(\hat{R}_X(X_i) - W_i' \hat{\gamma}) \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i) - W_i' \beta)(\hat{R}_X(X_i) - W_i' \gamma) + o_P(1). \end{aligned}$$

*Proof.* Using  $\sum_{i=1}^n W_i(\hat{R}_X(X_i) - W_i' \hat{\gamma}) = 0$ , we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i))(\hat{R}_X(X_i) - W_i' \hat{\gamma}) \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i) - W_i' \beta)(\hat{R}_X(X_i) - W_i' \hat{\gamma}) \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \hat{R}_X(X_i) - W_i' \beta)(\hat{R}_X(X_i) - W_i' \gamma) - R_n \end{aligned}$$

where

$$R_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho \widehat{R}_X(X_i) - W_i' \beta) W_i' (\widehat{\gamma} - \gamma) = R_{1n} + R_{2n}$$

with

$$\begin{aligned} R_{1n} &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i W_i' (\widehat{\gamma} - \gamma) \\ R_{2n} &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\rho R_X(X_i) - \rho \widehat{R}_X(X_i)) W_i' (\widehat{\gamma} - \gamma) \end{aligned}$$

By the same arguments as in the proof of Lemma 10, we have  $|R_{1n}| = o_P(1)$  and  $|R_{2n}| = o_P(1)$  and the desired result follows. Q.E.D.

### E.3.3 Proofs for Section 4.3

*Proof of Theorem 4.* As in the proof of Theorem 1,

$$\sqrt{n}(\widehat{\beta}_l - \beta_l) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - \beta_l W_{l,i})(W_{l,i} - W'_{-l,i} \widehat{\gamma}_l)}{\frac{1}{n} \sum_{i=1}^n (W_{l,i} - W'_{-l,i} \widehat{\gamma}_l)^2}$$

with

$$\widehat{\gamma} = \left( \sum_{i=1}^n W_{-l,i} W'_{-l,i} \right)^{-1} \left( \sum_{i=1}^n W_{-l,i} W_{l,i} \right).$$

By Lemmas 16 and 17, we have

$$\sqrt{n}(\widehat{\beta}_l - \beta_l) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - W_i' \beta)(W_{l,i} - W'_{-l,i} \gamma_l)}{\frac{1}{n} \sum_{i=1}^n (W_{l,i} - W'_{-l,i} \gamma_l)^2} + o_P(1). \quad (56)$$

Define  $Z_i := (Y_i, W_i)'$  for all  $i = 1, \dots, n$ ,

$$f(Z_i, Z_j) := (I(Y_j, Y_i) - W_i' \beta)(W_{l,i} - W'_{-l,i} \gamma_l)$$

for all  $i, j = 1, \dots, n$ , and  $h(Z_i, Z_j) := 2^{-1}(f(Z_i, Z_j) + f(Z_j, Z_i))$  for all  $i, j = 1, \dots, n$ . Since Assumption 2 implies  $E[h(Z_i, Z_j)^2] < \infty$ , we can argue as in the proof of Theorem 1 that

$$\frac{1}{n} \sum_{i=1}^n (\widehat{R}_Y(Y_i) - W_i' \beta)(W_{l,i} - W'_{-l,i} \gamma_l) = \frac{1}{n} \sum_{i=1}^n \left\{ h_{l,1}(W_i, Y_i) + h_{l,2}(Y_i) \right\} + o_P(n^{-1/2}) \quad (57)$$

where for  $z = (y, w)'$ , we define

$$\begin{aligned} E[f(Z_i, Z_j) \mid Z_i = z] &= (R_Y(y) - w' \beta)(w_l - w'_{-l} \gamma_l) = h_{l,1}(w, y), \\ E[f(Z_j, Z_i) \mid Z_i = z] &= E \left[ (I(y, Y) - W' \beta)(W_l - W'_{-l} \gamma_l) \right] = h_{l,2}(y). \end{aligned}$$

Combining (56) with (57) then implies the desired result by the same argument as that in the proof of Theorem 1. Q.E.D.

**Lemma 16.** Under Assumption 2, for each  $l = 1, \dots, p$ , we have  $\sigma_{\nu_l}^2 > 0$  and

$$n^{-1} \sum_{i=1}^n (W_{l,i} - W'_{-l,i} \hat{\gamma}_l)^2 \rightarrow_P E[(W_l - W'_{-l} \gamma_l)^2] = \sigma_{\nu_l}^2.$$

*Proof.* Since it follows from Assumption 2 that  $\sigma_{\nu_l}^2 > 0$ , the proof is similar to that of Lemma 9. Q.E.D.

**Lemma 17.** Under Assumption 2, for each  $l = 1, \dots, p$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{R}_Y(Y_i) - \beta_l W_{l,i})(W_{l,i} - W'_{-l,i} \hat{\gamma}_l) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{R}_Y(Y_i) - W'_i \beta)(W_{l,i} - W'_{-l,i} \gamma_l) + o_P(1).$$

*Proof.* Similar to that of Lemma 10. Q.E.D.

## E.4 Proofs for Appendix D

*Proof of Theorem 5.* Fix  $l = 1, \dots, p$ . As in the proof of Theorem 1,

$$\sqrt{n}(\hat{\beta}_l - \beta_l) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{R}_Y(Y_i) - \beta_l W_{l,i})(W_{l,i} - \hat{\tau}_l \hat{R}_X(X_i) - W'_{-l,i} \hat{\delta}_l)}{\frac{1}{n} \sum_{i=1}^n (W_{l,i} - \hat{\tau}_l \hat{R}_X(X_i) - W'_{-l,i} \hat{\delta}_l)^2},$$

where

$$\begin{pmatrix} \hat{\tau}_l \\ \hat{\delta}_l \end{pmatrix} = \left( \sum_{i=1}^n \begin{pmatrix} \hat{R}_X(X_i) \\ W_{-l,i} \end{pmatrix} \begin{pmatrix} \hat{R}_X(X_i) & W'_{-l,i} \end{pmatrix} \right)^{-1} \sum_{i=1}^n \begin{pmatrix} \hat{R}_X(X_i) \\ W_{-l,i} \end{pmatrix} W_{l,i}.$$

Further, again by the same arguments as those in the proof of Theorem 1,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{R}_Y(Y_i) - \beta_l W_{l,i})(W_{l,i} - \hat{\tau}_l \hat{R}_X(X_i) - W'_{-l,i} \hat{\delta}_l) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{R}_Y(Y_i) - \rho \hat{R}_X(X_i) - W'_i \beta)(W_{l,i} - \tau_l \hat{R}_X(X_i) - W'_{-l,i} \delta_l) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ h_{l,1}(X_i, W_i, Y_i) + h_{l,2}(X_i, Y_i) + h_{l,3}(X_i) \right\} + o_P(1) \end{aligned}$$

and

$$\frac{1}{n} \sum_{i=1}^n (W_{l,i} - \hat{\tau}_l \hat{R}_X(X_i) - W'_{-l,i} \hat{\delta}_l)^2 \rightarrow_P \sigma_{\nu_l}^2.$$

Combining these results with the central limit theorem and Slutsky's lemma gives the asserted claim. Q.E.D.

*Proof of Theorem 6.* Similar to that of Theorem 2, based on the same relationship as that between Theorem 5 and Theorem 1. Q.E.D.

*Proof of Theorem 7.* Similar to that of Theorem 3, based on the same relationship as that between Theorem 5 and Theorem 1. Q.E.D.