I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17591

Conditional Rank-Rank Regression

Victor Chernozhukov
Iván Fernández-Val
Jonas Meier
Aico Van Vuuren
Francis Vella

JANUARY 2025

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17591

Conditional Rank-Rank Regression

Victor Chernozhukov
MIT

Iván Fernández-Val
Boston University

Jonas Meier
Swiss National Bank

Aico Van Vuuren
University of Groningen and IZA

Francis Vella
Georgetown University and IZA

JANUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Conditional Rank-Rank Regression*

Rank-rank regression is commonly employed in economic research as a way of capturing the relationship between two economic variables. It frequently features in studies of intergenerational mobility as the resulting coefficient, capturing the rank correlation between the variables, is easy to interpret and measures overall persistence. However, in many applications it is common practice to include other covariates to account for differences in persistence levels between groups defined by the values of these covariates. In these instances the resulting coefficients can be difficult to interpret. We propose the conditional rank-rank regression, which uses conditional ranks instead of unconditional ranks, to measure average within-group income persistence. The difference between conditional and unconditional rank-rank regression coefficients can then be interpreted as a measure of between-group persistence. We develop a flexible estimation approach using distribution regression and establish a theoretical framework for large sample inference. An empirical study on intergenerational income mobility in Switzerland demonstrates the advantages of this approach. The study reveals stronger intergenerational persistence between fathers and sons compared to fathers and daughters, with the within-group persistence explaining 62% of the overall income persistence for sons and 52% for daughters. Smaller families and those with highly educated fathers exhibit greater persistence in economic status.

Corresponding author:
Francis Vella
Economics Department
Georgetown University
37th and O Streets, NW
Washington, DC 20057
USA

E-mail: fgv@georgetown.edu

* We thank Andrew Chesher, Toru Kitagawa, Patrick Kline, Michal Kolesar, Essie Maasoumi, Konrad Menzel, Ulrich Muller and seminar participants at Brown, BU, CUHK-SZ, Emory, Groningen, IESR, NYU, Princeton, UCL, Warwick and Oxford Workshop on Recent Advances in Panel and Network Data for helpful comments, and Matt Hong and Stella Hong for able research assistance.

1. Introduction

Rank-rank regressions (RRRs), in which the ranks of one variable, Y, are regressed on those of another, W, have been commonly employed to provide estimates of intergenerational mobility for a variety of economic outcomes. For example, Beller and Hout (2006) regressed a child's occupation rank on their father's occupation rank to measure intergenerational social mobility. Dahl and DeLeire (2008) and Chetty et al. (2014), among others, regressed child's income rank on father's income rank to measure intergenerational income mobility.[1] Adermon et al. (2018) regressed child's wealth rank on parent's and grandparent's wealth ranks to measure the role of inheritance in intergenerational wealth persistence. The use of RRRs has also extended beyond studies of intergenerational mobility. For example, Murphy and Weinhardt (2020) regressed a child's rank in a school grade on their rank in a previous grade to capture learning persistence. An appealing feature of the RRR as a measure of mobility is that its slope corresponds to the Spearman's rank correlation coefficient between the underlying variables Y and W.[2] This coefficient is a popular measure of dependence between variables that is invariant to monotone transformations (Spearman, 1904; Kendall, 1948).

The equivalence between RRR and rank correlation holds in the canonical RRR which features the rank of Y as the dependent variable and the rank of W as the exclusive independent variable. However, most empirical investigations which employ RRR also control for additional observed covariates X, as regressors. Chetverikov and Wilhelm (2023) observed that the coefficient of the rank of W in a RRR with covariates (RRRX) is difficult to interpret and does not necessarily lie in the interval $[-1, 1]$. Below we provide a simple example of intergenerational height mobility in two countries where the coefficient of the father's height rank is estimated to be greater than one when a country indicator is included as an additional covariate. Taken literally, this implies that moving the father's rank from zero to one would predict that the child's rank increases above one on average. This is clearly not credible. We also use this example to illustrate an additional problem with RRRs. In empirical investigations which feature a categorical covariate, it is common to conduct a subgroup analysis. That is, estimate separate RRR of marginal ranks for each group as defined by the categories of this covariate. We show that the coefficients of these RRRs also do not correspond to rank correlations and do not necessarily lie in the interval $[-1, 1]$. Moreover, they can be misleading as measures of within-group persistency. The underlying cause of these problems is that the independent variable in these regressions does not satisfy the properties of a rank within each group after partialling out the effect of the covariates.

We introduce the conditional rank-rank regression (CRRR) which does not suffer from these conceptual problems. CRRR regresses the rank of Y *conditional* on X on the rank of W *conditional* on X. Similar to the canonical RRR, the slope of the CRRR lies in the interval $[-1, 1]$ and has a

[1]Kitagawa et al. (2018) studied measurement error in RRRs in the context of intergenerational income mobility.

[2]Maasoumi et al. (2022) questioned the economic interpretation of the RRR as a measure of mobility due to the use of linear regression and proposed alternative measures based on nonparametric regression.

natural interpretation in terms of rank correlation between Y and W. Indeed, it corresponds to the Spearman's rank correlation between Y and W conditional on X, averaged over the distribution of X, which is a summary measure of within-group persistence. The CRRR is also suitable for subgroup analysis. If the categorical variable defining groups is included in X, the slope of the CRRR in each group has the interpretation of average conditional rank correlation in that group and lies in the interval $[-1, 1]$. We also show that if the conditional ranks of Y and W are constructed using different sets of covariates, the CRRR slope can still be interpreted as the average rank correlation conditional on the intersection of the two covariate sets.

The interpretation of the CRRR slope is different from the RRR slope. Assume, for example, that Y is child's income, W is father's income and X is a father's high school diploma indicator. The RRR slope without covariates is the correlation between the father's and child's income ranks where the ranks are relative to the entire income distribution. The CRRR slope is the rank correlation where the ranks are relative to the income distribution of those who have the same father's high school diploma indicator. The slope of the RRRX is the regression slope of the child's income rank on the father's income rank, where the ranks are relative to the entire income distribution and the father's rank is centered to have the same mean for both groups defined by the father's high school diploma indicator. The slope of this regression might be difficult to interpret as the centered rank is no longer a rank. We believe that the CRRR slope better reflects the relationship researchers are intending to capture when they control for covariates as it is closer to a ceteris paribus effect. Mathematically, the difference between CRRR and RRRX is the order in the application of the rank and covariate partialling out operators. RRRX obtain ranks first and partials out the covariates second, whereas CRRR reverses the order. The final outcome differs across procedures as the two operators do not commute due to the nonlinearity of the rank operator.

We provide an estimator of the CRRR coefficients based on distribution regression (DR). Like the estimator of the RRR coefficients, our estimator consists of two steps. In the RRR, the first step estimates the marginal ranks of Y and X using the empirical distribution, and the second step runs the linear regression of the estimated ranks of Y on the estimated ranks of W or computes the sample correlation between these ranks. Both versions of the second step produce numerically identical results if there are no ties in the observed values of Y and W. In the CRRR, the first step estimates the conditional ranks by running logit or probit DRs of Y on X and W on X at multiple values of Y and W to trace the entire conditional distributions. The second steps are identical to RRR, but the linear regression and correlation versions are no longer numerically identical even if there are no ties. They are, however, asymptotically equivalent. The CRRR estimator is computationally tractable, albeit somewhat more demanding than RRR.

We derive the asymptotic distribution of the CRRR estimator and provide feasible inference theory. Chetverikov and Wilhelm (2023) noted that standard inference methods for linear regression do not apply to the RRR estimator because both the independent and dependent variables, namely the estimated ranks, are generated. The same problem applies to CRRR. The theory for the RRR estimator was derived using U-statistic theory (Hoeffding, 1948) or the delta method (Ren and Sen, 1995) when Y and W are continuous. We employ the functional delta method approach to derive the theory because the CRRR estimator does not have a U-statistic structure. The application of the functional delta method to the CRRR estimator presents several differences with respect to RRR. The first ingredient of both approaches consists of writing the parameter of interest as a functional of inputs: the joint distribution of Y and W for RRR, or the conditional distributions of Y and W conditional on X and the joint distribution of Y, W and X, for CRRR. The CRRR functional is more complicated than the RRR functional and the inputs for CRRR live in more complex spaces than those for RRR. As a result, the Hadamard differentiability of the RRR functional established by Ren and Sen (1995) does not cover the CRRR functional. We establish the Hadamard differentiabilty of the CRRR functional in the relevant spaces.

The second ingredient required for the application of the functional delta method is to establish functional central limit theorems for the estimators of the inputs. For RRR, this follows from the now classical large sample theory of the empirical distribution function. A challenge for CRRR is that existing theory for DR estimators of conditional distributions exclude the tails. In particular, the available functional central limit theorems only hold for compact strict subsets of the support. We deal with this problem by imposing assumptions on the tail behavior of DR model that allows us to estimate the conditional distribution in the tails. We then obtain functional central limit theorems for DR estimators of conditional distributions over the entire support.

Combining these two ingredients we establish that the CRRR estimator follows a normal distribution around the CRRR slope in large samples via the delta method. The asymptotic variance has a complicated expression that might be difficult to estimate analytically. We develop the use of the exchangeable bootstrap to obtain standard errors and construct confidence intervals. Exchangeable bootstrap include the most common forms of bootstrap such as empirical, weighted and subsampling bootstrap as special cases. We establish its validity in large samples from the functional delta method for the bootstrap. Like Hoeffding (1948) and Ren and Sen (1995), our theory covers the case where Y and W are continuous. The theory can be extended to noncontinuous variables following the analysis of Chetverikov and Wilhelm (2023) for the RRR estimator. We leave this extension to future research.

We apply the CRRR estimator to analyze intergenerational income mobility in Switzerland using the Economic Well-Being of the Working and Retirement Age Population Data (WiSiER) from 1986 to 2016 for 11 cantons. This dataset contains rich information on socioeconomic, demographic and other variables merged from tax records, social insurance, unemployment records and surveys, and can be linked for fathers and children. We uncover a gender gap in intergenerational income mobility as the persistence between fathers and sons is stronger than between fathers and daughters, both with and without controlling for covariates. We also find that about 62% and 52% of the overall (unconditional) income persistence is explained by the within-group income persistence for sons and daughters, respectively; where groups are defined by child's and father's marital status, Swiss citizenship, high school graduation, experience, number of children and canton and year fixed effects. We also provide evidence supporting greater persistence for fathers with higher education and fathers with only one child. These results uncover the substantial role of both within-group and between-group persistence in explaining the intergenerational transmission of income.

Our methodology complements related methodological developments in the econometric and statistical literature. While Chetverikov and Wilhelm (2023) provides the inferential theory for RRR and RRRX with marginal ranks, their approach does not apply to conditional ranks employed in CRRR, as we explained earlier. Another new development is Lei (2024) which examines causal underpinnings of the marginal rank regressions. More closely related to our work, Liu et al. (2018) introduced a covariate-adjusted Spearman coefficient based on the probability scale residuals of Li and Shepherd (2012) and Shepherd et al. (2016), which we further discuss in Section 2. They proposed a modelling strategy and an estimator based on a monotone transformation of a location-shift model, which is a special case of the distribution regression model (Chernozhukov et al., 2013).[3] The DR approach is more flexible and comprehensive, in that it can approximate the true conditional distribution function arbitrarily well by considering rich sets of basis functions with respect to the covariates. This is not generally possible with transformations of location models.[4] Gijbels et al. (2011) and Veraverbeke et al. (2011) developed estimators of conditional measures of association using copulas, relying on the representation of these measures in terms of the conditional copula for continuous outcomes. They derived distribution theory for estimators of conditional versions of Kendall's tau and Blomqvist's beta (Blomqvist, 1950) with scalar covariates. In contrast, our modelling and estimators are based on conditional distributions and apply to multivariate covariates.

Outline. The rest of the paper is organized as follows. Section 2 introduces the CRRR and contrasts it with the canonical RRR and RRRX. Section 3 illustrates the problems with the RRR with covariates via a simple conceptual example. Section 4 describes the estimation procedure based on DR and a bootstrap algorithm to make inference. Section 5 discusses an application examining the relationship between fathers' and their children's labor income using Swiss data, while Section 6 provides asymptotic theory. Additional theoretical and numerical simulation results, and proofs are reported in the Appendix.

[3]Liu et al. (2018) did not develop inference theory for their estimator in the case where Y and W are continuous, although they conjectured that it is possible; see Section 5 ibid.

[4]A transformation of a location-shift model takes the form $H(Y) = b(X)'\beta + \epsilon$, where ϵ is independent of X and has a known distribution, $b(X)$ is a basis of functions of X, and H is an unknown monotone transformation function. The model is more flexible than just a location model, but it does not allow the covariates to affect the distribution of $H(Y)$ other than through its location. No matter how rich the basis functions $b(X)$ are, the model is not guaranteed to cover the true conditional distribution, even in the limit where the dimension of $b(X)$ grows large.

2. Conditional Rank-Rank Regression

Let (Y, W) be a bivariate random variable with joint distribution $F_{Y,W}$ and marginal distributions F_Y and F_W for Y and W, respectively. For example, Y is child's income and W is father's income. We assume that Y and W are continuous.

2.1. **Canonical RRR.**

We start by reviewing the canonical rank-rank regression (RRR). Let $\tilde{U} := F_Y(Y)$ and $\tilde{V} := F_W(W)$ denote the ranks of Y and W. By continuity of Y and W, ranks are uniformly distributed, $\tilde{U} \sim U(0,1)$ and $\tilde{V} \sim U(0,1)$. The RRR of Y on W is defined as the correlation between $\tilde{U}$ and $\tilde{V}$ or the slope of the linear regression of $\tilde{U}$ on $\tilde{V}$ (or vice versa):

$$\rho := \mathrm{Cor}(\tilde{U}, \tilde{V}) = \frac{\mathrm{Cov}(\tilde{U}, \tilde{V})}{\mathrm{Var}(\tilde{U})} = \frac{\mathrm{Cov}(\tilde{U}, \tilde{V})}{\mathrm{Var}(\tilde{V})} = 12\,\mathrm{E}[(\tilde{U} - .5)(\tilde{V} - .5)],$$

where all the equalities follow from the uniform distribution of $\tilde{U}$ and $\tilde{V}$. In statistics this correlation measure is the celebrated Spearman's rank correlation between Y and W, and is widely used to measure dependence between variables. It is invariant to rescaling and all increasing monotone transformations of the variables, and has gained prominence for that reason. The rank correlation has become popular in economics in studies of income and wealth mobility due to its interpretability as a measure of persistence and scale-free nature.

2.2. **Conditional RRR.**

We introduce now the conditional rank-rank regression (CRRR). Let X denote a vector of covariates related to Y and W including, for example, child's and father's education, age, marital status and nationality. Let $F_{Y|X}$ and $F_{W|X}$ denote the distributions of Y and W conditional on X. Then, $U := F_{Y|X}(Y \mid X)$ and $V := F_{W|X}(W \mid X)$ are the conditional ranks of Y and W, where conditioning is on X. For example, U and V would be child's and father's income ranks among families with the same composition in terms of covariates. By continuity of Y and W, the conditional ranks follow the uniform distribution, conditional on X:

$$U \mid X \sim U(0,1) \text{ and } V \mid X \sim U(0,1),$$

and also unconditionally. This implies the constant variance property,

$$\mathrm{Var}(V) = \mathrm{Var}(U) = \mathrm{Var}(V \mid X) = \mathrm{Var}(U \mid X) = 1/12$$

and the constant mean property,

$$\mathrm{E}V = \mathrm{E}U = \mathrm{E}(V \mid X) = \mathrm{E}(U \mid X) = .5.$$

Note that both U and V are marginally independent of X , but not necessarily jointly independent so the correlation between U and V can depend on X.[5]

[5]That is, $U \perp\!\!\!\perp X$ and $V \perp\!\!\!\perp X$, but generally $(U, V) \not\perp\!\!\!\perp X$, where $\perp\!\!\!\perp$ denotes stochastic independence.

The CRRR of Y on W given X is defined as either the correlation between U and V or the slope of the linear regression of U on V (or vice versa):

$$\rho_C = \mathrm{Cor}(U, V) = \frac{\mathrm{Cov}(U, V)}{\mathrm{Var}(V)} = \frac{\mathrm{Cov}(U, V)}{\mathrm{Var}(U)}. \tag{2.1}$$

CRRR is the average conditional correlation between conditional ranks:

$$\rho_C = \mathrm{E}[\rho_{Y,W|X}], \quad \rho_{Y,W|X} := \mathrm{Cor}(U, V \mid X), \tag{2.2}$$

where $\rho_{Y,W|X}$ denotes the conditional Spearman's rank correlation between Y and W conditional on X, which is equal to $\mathrm{Cor}(U, V \mid X)$ by definition. Equation (2.2) follows from $\mathrm{Cov}(U, V) = \mathrm{E}[\mathrm{Cov}(U, V \mid X)]$ by the law of total covariance since $\mathrm{Cov}[\mathrm{E}(U \mid X), \mathrm{E}(V \mid X)] = 0$; moreover, the conditional variance of U and V is equal to the unconditional variance. In summary, CRRR is the average Spearman's rank correlation between Y and W conditional on X, averaged over the distribution of X, wich is a summary measure of within-group persistence.

By the properties of U and V, the CRRR can also be represented as the rescaled covariance of conditional ranks:

$$\rho_C = 12\,\mathrm{E}[(U - .5)(V - .5)], \tag{2.3}$$

a formula convenient for estimation.

Finally, we note that correlation of conditional ranks is generally not equal to correlation of marginal (unconditional) ranks:

$$\rho_C \neq \rho$$

but the two agree under independence from X, namely $\rho_C = \rho$ if $Y \perp\!\!\!\perp X$ and $W \perp\!\!\!\perp X$, because in that case $U = \tilde{U}$ and $V = \tilde{V}$.

In the context of the income mobility application, ρ_C measures within-group income persistence and ρ measures overall income persistence, encompassing both within-group and between-group persistence. The between-group persistence can then be defined as the difference between the marginal rank and conditional rank correlations:

$$\text{Between-group persistence} = \rho - \rho_C.$$

Assume, for example, that the covariates X capture family characteristics such as size or parental education. The difference between the two measures can be explained as follows: The within-group or unexplained persistence ρ_C captures the extent to which father's income rank facilitates child's income rank among families with the same observable characteristics. In other words, it measures the influence of father's income on child's income, where the variation in father's and child's incomes comes from unobserved characteristics such as family status, ability and the extent of social or professional networks. On the other hand, the between-group measure $\rho - \rho_C$ aims to capture the contribution of observed characteristics to income persistence.

We can further decompose the between-group persistence using the total law of covariance:

$$\rho - \rho_C = 12\,\mathrm{Cov}[\mathrm{E}(\tilde{U} \mid X), \mathrm{E}(\tilde{V} \mid X)] + 12\mathrm{E}[\mathrm{Cov}(\tilde{U}, \tilde{V} \mid X) - \mathrm{Cov}(U, V \mid X)],$$

where the first component is the covariance of conditional means of marginal ranks, and the second component is the average conditional covariance of marginal ranks net of the average within-group inequality.

2.3. **Rank-rank regression with covariates (RRRX).** CRRR is different from RRR with covariates X (RRRX) where X is included additively (or non-additively) in the regression of marginal ranks, $\tilde{U}$ on $\tilde{V}$. We believe that our proposal is a more natural and adequate way to incorporate covariates. In fact, RRRX with additive covariates is no longer related to a rank correlation nor has to lie in the interval $[-1, 1]$. RRRX is also more difficult to interpret as it does not correspond to a meaningful measure of within-group persistence. Making RRRX more flexible by including interactions between X and $\tilde{V}$ does not mitigate any of these problems. In fact, making RRRX fully nonparametric also does not alleviate the problem. We show in the next section that even in the simplest case where X is binary, the nonparametric RRRX does not capture meaningful economic quantities. When X is discrete, the nonparametric approach (tabulating unconditional rank correlation by subgroups) does not either.

In what follows, we systematically explain the current approaches to RRRX and contrast these with the CRRR approach. We use the intergenerational income application to give context to the discussion.

Example 1 (RRR vs RRRX vs CRRR)**.** Let Y be child's income, W be father's income and X be an indicator for father's high school diploma. In this case, the marginal ranks $\tilde{U}$ and $\tilde{V}$ are relative to the distribution of income in the entire population that includes fathers with and without high school diploma, whereas the conditional ranks U and V are relative to the distribution of income of those with the same father's high school diploma status. RRR measures the correlation between the marginal ranks, whereas CRRR measures the average correlation between the conditional ranks, that is CRRR first obtains the rank correlation separately for fathers with and without high school diploma and then averages these correlations weighted by the proportions of each type in the population. CRRR therefore can be interpreted as a within-group or *ceteris paribus* effect, where the families are ranked and compared with families where the father's high school diploma status is held constant. The slope of the RRRX with covariates does not have a natural interpretation in terms of intergenerational mobility. It measures the coefficient in the regression of child's marginal rank on father's marginal rank, where the father's marginal rank is recentered to have the same mean for fathers with and without high school diploma. This slope does not have an interpretation as a within-group persistence. Moreover, it is not a rank correlation and can lie outside the interval $[-1, 1]$, because the recentered father's marginal rank does not have the properties of a rank. In particular it no longer follows a uniform distribution. ∎

2.4. **Subgroup Analysis.** When X is discrete, it is common to run RRRs separately for each value of X instead of including X as an additive control. For example, Abramitzky et al. (2021) run separate RRR of child's income on father's income by father's immigration status. The slopes of these regressions cannot be interpreted in terms of rank correlations or even as conditional correlations between the marginal ranks. To see this, note that the slope of the regression of $\tilde{U}$ on $\tilde{V}$ conditional on $X = x$, is not equal to conditional correlation of $\tilde{U}$ and $\tilde{V}$:

$$\frac{\operatorname{Cov}(\tilde{U}, \tilde{V} \mid X = x)}{\operatorname{Var}(\tilde{V} \mid X = x)} \neq \frac{\operatorname{Cov}(\tilde{U}, \tilde{V} \mid X = x)}{\sqrt{\operatorname{Var}(\tilde{V} \mid X = x) \operatorname{Var}(\tilde{U} \mid X = x)}},$$

because marginal ranks have different conditional distributions, i.e. $\tilde{U} \overset{d}{\not\sim} \tilde{V} \mid X = x$, in general. The slope therefore does not generally correspond with the conditional correlation of the marginal ranks conditional nor the conditional rank correlation between Y and W. We give an example in Section 3 where this slope is greater than one.

Consider now the CRRR. Assume we are interested in conducting a subgroup analysis of intergenerational mobility with respect to father's high school diploma or immigration status. Let $X_1 \subseteq X$ be a set of variables that define the subpopulation of interest such as an indicator for high school diploma and/or Swiss nationality. Then, the CRRR slope conditional on $X_1 = x_1$ is:[6]

$$\rho_C(x_1) = \frac{\operatorname{Cov}(U, V \mid X_1 = x_1)}{\operatorname{Var}(V \mid X_1 = x_1)} = \operatorname{E}\left[\frac{\operatorname{Cov}(U, V \mid X)}{\sqrt{\operatorname{Var}(V \mid X) \operatorname{Var}(U \mid X)}} \mid X_1 = x_1\right] = \operatorname{E}[\rho_{Y,W|X} \mid X_1 = x_1]. \tag{2.4}$$

Hence, the CRRR slope for the subgroup defined by $X_1 = x_1$ corresponds to the average conditional rank correlation between Y and W, where the average is taken with respect to the distribution of X conditional on $X_1 = x_1$. This allow us, for example, to measure intergenerational mobility separately for families with fathers with and without high school diploma.[7]

2.5. **Different Sets of Covariates.** There are applications where the researcher might want to use different sets of covariates to obtain the conditional ranks U and V. In the intergenerational mobility application, for example, we might not want to control for son's education to obtain the father's income rank. In this case the CRRR slope still corresponds to an average correlation between the ranks. To see this, let $U = F_{Y|X_1}(Y \mid X_1)$ and $V = F_{Y|X_2}(Y \mid X_2)$ with $X_1 \neq X_2$ and $X = X_1 \cap X_2$,

[6]Indeed, by the law of total covariance with respect to X and uniformity of U and V conditional on X,

$$\operatorname{Cov}(U, V \mid X_1) = \operatorname{E}[\operatorname{Cov}(U, V \mid X) \mid X_1] + \operatorname{Cov}[\operatorname{E}(U \mid X), \operatorname{E}(V \mid X) \mid X_1) = \operatorname{E}[\operatorname{Cov}(U, V \mid X) \mid X_1],$$

and

$$\operatorname{Var}(V \mid X_1) = \operatorname{Var}(V \mid X) = \operatorname{Var}(U \mid X),$$

almost surely.

[7]If $X_1 \not\subseteq X$, , the slope no longer has an interpretation as average conditional rank correlation because $V \not\sim U \mid X_1$ in general.

then:

$$\rho_C = \frac{\mathrm{Cov}(U,V)}{\mathrm{Var}(V)} = \mathrm{E}\left[\frac{\mathrm{Cov}(U,V \mid X)}{\sqrt{\mathrm{Var}(V \mid X)\,\mathrm{Var}(U \mid X)}}\right],$$

where we use the law of total covariance with respect to X, $U \perp\!\!\!\perp X$, $V \perp\!\!\!\perp X$ and iterated expectations. The CRRR slope therefore corresponds to the correlation between the ranks U and V conditional on the common covariates X, averaged over the distribution of X. Note, however, that ρ_C in this case does not correspond to an average conditional rank correlation between Y and W. The source of the difference is that $U \neq F_{Y|X}(Y \mid X)$ and $V \neq F_{W|X}(W \mid X)$ in general.[8] One exception occurs when Y is independent of the components of X_2 not included in X_1 conditional on X_1, and W is independent of the components of X_1 not included in X_2 conditional on X_2. In that case,

$$\rho_C = \mathrm{E}\left[\frac{\mathrm{Cov}(U,V \mid \bar{X})}{\sqrt{\mathrm{Var}(V \mid \bar{X})\,\mathrm{Var}(U \mid \bar{X})}}\right] = \mathrm{E}[\rho_{Y,W|\bar{X}}],$$

where $\bar{X} = X_1 \cup X_2$. This result follows by the law of total covariance with respect to $\bar{X}$ and uniformity of V and U conditional on $\bar{X}$.[9]

An interesting example occurs when $X_1 = \emptyset$ and $X_2 = X$. In this case, $U = \tilde{U}$ and ρ_C is the correlation between the marginal ranks of Y and the conditional ranks of W. Unlike the RRR, the inclusion of covariates in this CRRR does not affect the coefficient of V because $V \perp\!\!\!\perp X$, and can be used to perform a variance decomposition of $\tilde{U}$. Let,

$$\tilde{U} = \rho_C V + X'\beta_C + \varepsilon, \quad \mathrm{E}[(V;X)\varepsilon] = 0,$$

be the extended CRRR with covariates, where the first term of X is a constant. Then,

$$\mathrm{Var}(\tilde{U}) = \rho_C^2 \mathrm{Var}(V) + \mathrm{Var}(X'\beta_C) + \mathrm{Var}(\varepsilon),$$

where the first two terms of the right-hand-side correspond to the contribution of V and X to the variance of $\tilde{U}$, and the third terms to the unobserved component. Indeed, ρ_C^2 measures the fraction of the variance of $\tilde{U}$ explained by V since $\mathrm{Var}(\tilde{U}) = \mathrm{Var}(V)$.

2.6. **Properties of CRRR.** We conclude this section by gathering the properties of the CRRR slope in the following lemma.

Lemma 2.1 (CRRR Properties)**.** *Assume that Y and W are continuous random variables, X is a vector of covariates, $U = F_{Y|X}(Y \mid X)$ and $V = F_{W|X}(W \mid X)$. Then,* (1) *The CRRR slope, ρ_C, has the representations given in* (2.1) *and* (2.3)*.* (2) *The slope ρ_C is the expected conditional Spearman rank correlation between Y and W: $\rho_C = \mathrm{E}[\rho_{Y,W|X}]$.* (3) *Group analysis: if $X_1 \subseteq X$, then* (2.4) *holds. Therefore, $\rho_C(x_1)$*

[8] This rank correlation can be obtained by constructing the conditional ranks as $U = F_{Y|X}(Y \mid X)$ and $V = F_{W|X}(W \mid X)$.

[9] Note that if $V \mid X_1 \sim U(0,1)$ and $V \perp\!\!\!\perp X_2 \mid X_1$, then $V \mid X \sim U(0,1)$.

is the average conditional rank correlation between Y and W in the group defined by $X_1 = x_1$. (4) Let $U_1 = F_{Y|X_1}(Y \mid X_1)$ and $V_2 = F_{Y|X_2}(Y \mid X_2)$ with $X_1 \neq X_2$ and $X = X_1 \cap X_2$, then

$$\rho_C = \mathrm{Cor}(U_1, V_2) = \mathrm{E}[\mathrm{Cor}(U_1, V_2 \mid X)],$$

that is ρ_C is the average conditional correlation between U_1 and V_2 given the set of common covariates X.

Comment 2.1. This lemma simply records the observations given above. It is useful to connect here to Liu et al. (2018) who introduced the covariate-adjusted Spearman correlation coefficient as the correlation between the probability scale residuals of Y and W. These residuals are defined as $r(Y, F_{Y|X})$ and $r(W, F_{W|X})$, where $r(r, F_{R|X}) = F_{R|X}(r \mid X) + F_{R|X}(r- \mid X) - 1$ and $F_{R|X}(r- \mid x) = \lim_{u \nearrow r} F_{R|X}(u \mid x)$, for $R \in \{Y, W\}$. In the case where Y and W are continuous, the probability scale residuals are affine transformations of the conditional ranks because $F_{R|X}(r- \mid x) = F_{R|X}(r \mid x)$, e.g., $r(Y, F_{Y|X}) = 2U - 1$, and the covariate-adjusted Spearman correlation equals to the CRRR slope. The properties in Lemma 2.1(2) and (3) then follow from results in Liu et al. (2018) when Y and W are continuous. The conceptual difference is that our definition and derivations are based on the characterization of the Spearman correlation as the correlation between ranks or grade correlation (Kruskal, 1958), whereas theirs are based on the characterization of the Spearman correlation in terms of concordance-discordance probabilities.

3. CRRR vs RRR: A Conceptual Example

We now compare CRRR with different versions of RRR in a simple conceptual example where X is binary. This example is convenient because with a binary covariate there is no concern that the difference between the methods is driven by particular modeling strategies to specify various regression functions.

Let Y be daughter's height (in cm), W be father's height (in cm) and X be a country indicator, say $X = 0$ for the Netherlands and $X = 1$ for Ireland. Conditional on X, Y and W follow a bivariate normal distribution with mean parameters that may depend on the value of X and constant covariance matrix. More specifically,

$$\begin{pmatrix} Y \\ W \end{pmatrix} \mid X = x \sim N_2\left(\begin{pmatrix} 165 \\ 180 - \delta x \end{pmatrix}, 4^2 \begin{pmatrix} 1 & .6 \\ .6 & 1 \end{pmatrix} \right), \tag{3.1}$$

where $\mathrm{P}(X = 0) = \mathrm{P}(X = 1) = 1/2$. For example, δ can be a negative country shock such as the Irish Famine that affects the father's height in Ireland, but not in the Netherlands.

We consider two cases depending on the extent of the effect of the shock as measured by δ:

- No shock: $\delta = 0$.
- Negative shock: $\delta = 12$.

Table 1 compares measures of overall and within-country intergenerational height persistence based on rank correlation with the estimands of RRR, CRRR and two versions of RRRX. Thus, $\rho_{Y,W}$

and $\bar{\rho}_{Y,W|X}$ are the Spearman's rank correlation coefficient between Y and W and the expected Spearman's rank correlation between Y and W conditional on X; RRR is the RRR slope; CRRR is the CRRR slope; RRRX-A is the slope of the RRRX, that is, β_1 in

$$\tilde{U} = \beta_0 + \beta_1 \tilde{V} + \beta_2 X + \epsilon, \quad \mathrm{E}[\epsilon] = \mathrm{E}[\tilde{V}\epsilon] = \mathrm{E}[X\epsilon] = 0;$$

and RRRX-I is the average slope of the RRRs run separately by the values of X, that is, β_1 in

$$\tilde{U} = \beta_0 + \beta_1 \tilde{V} + \beta_2 [X - .5] + \beta_3 [X - .5]\tilde{V} + \epsilon, \quad \mathrm{E}[\epsilon] = \mathrm{E}[\tilde{V}\epsilon] = \mathrm{E}[X\epsilon] = \mathrm{E}[X\tilde{V}\epsilon] = 0$$

Table 1. Mobility Measures and Estimands

	Overall		Within-Country			
	True	Estimand	True	Estimand		
	$\rho_{Y,W}$	RRR	$\bar{\rho}_{Y,W\|X}$	RRRX-A	RRRX-I	**CRRR**
$\delta = 0$	0.58	0.58	0.58	0.58	0.58	**0.58**
$\delta = 12$	0.32	0.32	0.58	1.07	1.07	**0.58**

Notes: based on 2,000,000 simulations.

We find that all the methods give the same answer when $\delta = 0$, that is when the joint distribution of daughter's and father's heights is the same in both countries.[10] When $\delta = 12$, RRR gives the overall mobility $\rho_{Y,W}$, whereas CRRR gives the within-country $\bar{\rho}_{Y,W|X}$. Both forms of RRRX produce measures greater than one, which do not correspond to any rank correlation and might be difficult to interpret. Whether RRR or CRRR is the right measure depends on the application. In this case, CRRR measures average intergenerational mobility within each country whereas RRR measures intergenerational mobility pooling the two countries. They would lead to different conclusions about the effect of the Irish Famine. According to RRR, the famine reduces overall height persistence, whereas it does not have any effect within each country according to CRRR. Both versions of RRRX lead to the opposite conclusion that the famine increases height persistence. This conclusion does not correspond to the change in either overall or within-country mobility.

Table 2 compares the subgroup analysis based on CRRR and RRR. More specifically, we compare the conditional Spearman's rank correlation, CRRR slope and RRR slope for each value of X. CRRR produces measures that are invariant to both X and δ, which correspond to the rank correlations between Y and W conditional on X. The RRR slopes are the same as the CRRR slopes when X is irrelevant. RRR, however, delivers different slopes both across values of X when $\delta = 12$, and also across values of δ. The RRR slopes are greater than one when $\delta = 12$, confirming that they do not correspond to correlations and making them hard to interpret.

[10] Up to numerical error, all the slopes are equal to the rank correlation of the bivariate normal with correlation $c = .6$, $\rho_S(Y, W) = 6 \arcsin(c/2)/\pi = .58$ (Cramér, 1999).

TABLE 2. Subgroup Analysis by Country using RRRX and CRRR

	True		Estimand			
	$\rho_{Y,W\mid X}$		RRR		**CRRR**	
	$X=0$	$X=1$	$X=0$	$X=1$	$X=0$	$X=1$
$\delta=0$	0.58	0.58	0.58	0.58	**0.58**	**0.58**
$\delta=12$	0.58	0.58	1.06	1.07	**0.58**	**0.58**

Notes: based on 2,000,000 simulations.

To better understand the source of the differences between CRRR and RRR, we consider a finite sample example based on 20 observations of the design (3.1), 10 with $X=0$ and 10 with $X=1$, and with $\delta=12$. Table 3 shows the observations of Y and W, together with the conditional ranks, U and V, the marginal ranks, $\tilde{U}$ and $\tilde{V}$, and the residualized marginal rank of W after partialling out the effect of the covariate X, $\tilde{V}_r$. All the ranks are expressed in per cent and the residualized rank are recentered at .5 to have the same mean as the other ranks. Note that CRRR is the slope of the regression of U on V, RRR is the slope of the regression of $\tilde{U}$ on $\tilde{V}$, and RRRX-A is the slope of the regression of $\tilde{U}$ on $\tilde{V}_r$. Here we can see that the main source of the difference between RRR and CRRR in this case arises from the father's ranks $\tilde{V}$ and V. Thus, while the daughter's ranks U and $\tilde{U}$ are similar, the marginal ranks $\tilde{V}$ are relatively larger than the conditional ranks $\tilde{V}$ in the Netherlands and smaller in Ireland due to the location change in Ireland. This results in the RRR slope being smaller than the CRRR slope. Netting out the country effect from the father's marginal ranks brings them generally closer to the father's conditional ranks, but the residualized ranks are no longer ranks in that the are not uniformly distributed conditionally or unconditionally. As a result, the RRRX-A slope does not have an interpretation in terms of rank correlation and might yield values greater than one in absolute value. In other words, the difference between CRRR and RRRX-A is the order in the application of the rank and partialling out operators. CRRR computes the ranks after partialling out the effect of X, whereas RRRX-A reverses the order. They deliver different results because these operators do not commute due to the nonlinearity of the rank operator.

4. DISTRIBUTION REGRESSION ESTIMATOR OF CRRR

4.1. **DR Model for Conditional Distributions.** For estimation purposes, it is convenient to model the conditional distributions $F_{Y\mid X}$ and $F_{W\mid X}$ using the distribution regression (DR) model:

$$F_{R\mid X}(r\mid x)=\Lambda(x'\beta_R(r)),\quad R\in\{Y,W\},\quad r\in\mathcal{R},$$

TABLE 3. Finite Sample Example based on (3.1) with $\delta = 12$

X	Y	$U(\%)$	$\tilde{U}(\%)$	W	$V(\%)$	$\tilde{V}(\%)$	$\tilde{V}_r(\%)$
0	162	10	20	173	10	55	28
0	162	20	25	176	40	70	42
0	164	30	30	182	100	100	73
0	164	40	35	174	20	60	32
0	164	50	45	177	60	80	52
0	165	60	60	177	50	75	48
0	165	70	70	175	30	65	38
0	165	80	75	179	80	90	62
0	168	90	85	181	90	95	68
0	171	100	100	178	70	85	57
1	156	10	5	160	10	5	28
1	157	20	10	169	70	35	57
1	162	30	15	163	20	10	32
1	164	40	40	165	50	25	48
1	164	50	50	164	40	20	43
1	165	60	55	164	30	15	38
1	165	70	65	171	100	50	73
1	167	80	80	170	80	40	62
1	168	90	90	167	60	30	53
1	170	100	95	170	90	45	68

Notes: $\delta = 12$ and $\tilde{V}_r = .5 + \tilde{V} - \mathrm{E}[\tilde{V} \mid X]$.

where Λ is the standard normal or logistic distribution, $\mathcal{R}$ is the support of R and the first component of x is a constant. The specification can be made more flexible by replacing x by a vector of transformations of x with good approximating properties.

As the data to estimate the conditional distribution function at the tails are sparse, it is necessary to impose some structure. We assume that the conditional distribution far in the tails can be extrapolated from the conditional distribution not too far in the tails.[11] We formalize this approach by imposing restrictions on the coefficient of the DR model in the tails.

Let $\bar{\mathcal{R}}$ be a compact strict subset of $\mathcal{R}$, for $\mathcal{R} \in \{\mathcal{Y}, \mathcal{W}\}$. Then, we assume:

$$F_{R|X}(r \mid x) = \Lambda((r - \bar{r})\alpha_R(\bar{r}) + x'\beta_R(\bar{r})), \quad R \in \{Y, W\}, \quad r \in \mathcal{R} \setminus \bar{\mathcal{R}},$$

[11]This is in line with approaches used in extreme value theory that impose restrictions on the tail behavior allowing similar extrapolations. For example, see Embrechts et al. (1997) for a broad reference on the theory of extremes and Chernozhukov (2005) or Chernozhukov and Fernández-Val (2011) for similar approaches in the context of extremal quantile regression.

where $\bar{r} := \arg\min_{r' \in \bar{\mathcal{R}}} |r - r'|$ and $\alpha_R(\bar{r}) > 0$. That is, we postulate that the random variable R behaves in the tails like a random variable with distribution Λ, after subtracting the location shift $x'\beta_R(\bar{r})$ and dividing by the scale $\alpha_R(\bar{r})$, which are different at the upper and lower tails. Thus, the DR coefficient is restricted in the tails by:

$$\beta_{R,1}(r) = \beta_{R,1}(\bar{r}) + (r - \bar{r})\alpha_R(\bar{r}), \quad \beta_{R,-1}(r) = \beta_{R,-1}(\bar{r}), \quad R \in \{Y, W\}, \quad r \in \mathcal{R} \setminus \bar{\mathcal{R}},$$

where $\beta_R(r)$ is partitioned into $(\beta_{R,1}(r), \beta_{R,-1}(r)')'$ where $\beta_{R,1}(r)$ is the intercept and $\beta_{R,-1}(r)$ are the slope components. That is, $r \mapsto \beta_{R,1}(r)$ is a linear function and $r \mapsto \beta_{R,-1}(r)$ is constant on $\mathcal{R} \setminus \bar{\mathcal{R}}$.

Under the DR model, the conditional ranks can be expressed as the following functionals of the parameters:

$$U = \Lambda(X'\beta_Y(Y)), \quad V = \Lambda(X'\beta_W(W)).$$

4.2. **Estimation.** We provide several estimators of the CRRR slope based on the different representations of ρ_C in (2.1) and (2.3). This section presents correlation-based and fully-restricted estimators. Regression-based estimators are given in Appendix A. We recommend the use of at least the correlation-based and fully-restricted estimators. The fully-restricted estimator, based on (2.3), uses all the information available and is the simplest to compute, but it might be sensitive to misspecification of the model for the conditional distributions. In particular, it can deliver estimates outside the interval $[-1, 1]$ under misspecification. The correlation-based estimator is more robust in the sense that it is the only estimator that guarantees estimates in the interval $[-1, 1]$ under misspecification.[12] We show in Appendix A that the correlation-based estimator is asymptotically equivalent to the average of the regression-based and reversed regression-based estimators.

Let $\{Z_i := (Y_i, W_i, X_i)\}_{i=1}^n$ be a random sample of $Z := (Y, W, X)$. The following algorithms describe the estimators of ρ_C. All of them are based on DR.

Algorithm 1 (Correlation-based and Fully-Restricted Estimators)**.** Let $d_x := \dim X$, $\mathcal{R}_n$ denote the set containing the observed values of R and $\bar{\mathcal{R}}_n = \mathcal{R}_n \cap \bar{\mathcal{R}}$, for $R \in \{Y, W\}$.

(1) Estimate $\beta_R(r)$ at $r \in \bar{\mathcal{R}}_n$ by DR, that is,

$$\widehat{\beta}_R(r) \in \operatorname{argmax}_{b \in \mathbb{R}^{d_x}} \sum_{i=1}^{n} \left[1(R_i \leqslant r) \log \Lambda(X_i'b) + 1(R_i > r) \log \Lambda(-X_i'b) \right].$$

(2) Estimate $\beta_R(r)$ at $r \in \mathcal{R}_n \setminus \bar{\mathcal{R}}_n$ by restricted DR, that is,

$$\widehat{\beta}_R(r) = (r - \bar{r})\widehat{\alpha}_R(\bar{r}) + x'\widehat{\beta}_R(\bar{r}),$$

[12] While we impose correct specification of the DR model for the conditional distributions, the derivation of the theoretical results does not rely fundamentally on correct specification. We conjecture that the probability limit of the correlation-based estimator still has an interpretation as correlation of pseudo-ranks under misspecification, but leave the formal analysis to future research.

where $\bar{r} := \arg\min_{r' \in \bar{\mathcal{R}}_n} |r - r'|$ and

$$\widehat{\alpha}_R(\bar{r}) \in \operatorname{argmax}_{a \in \mathbb{R}} \sum_{i=1}^{n} \Big[1(R_i \leqslant r_0) \log \Lambda((r_0 - \bar{r})a + X_i'\widehat{\beta}_R(\bar{r})) + 1(R_i > r_0) \log \Lambda(-(r_0 - \bar{r})a - X_i'\widehat{\beta}_R(\bar{r})) \Big],$$

and $r_0 \in \mathcal{R}_n \setminus \bar{\mathcal{R}}_n$ is such that (i) there are at least m observations between $\bar{r}$ and r_0, and greater than r_0 if $r_0 > \bar{r}$ (upper tail) or less than r_0 if $r_0 < \bar{r}$ (lower tail), and (ii) $\widehat{\alpha}_R(\bar{r}) > 0$.

(3) Obtain plug-in estimators of the conditional ranks

$$\widehat{U}_i = \Lambda(X_i'\widehat{\beta}_Y(Y_i)), \quad \widehat{V}_i = \Lambda(X_i'\widehat{\beta}_W(W_i)).$$

(4) Estimate ρ_C as either (a) the sample correlation between $\widehat{U}_i$ and $\widehat{V}_i$, that is

$$\widehat{\rho}_C = \frac{\sum_{i=1}^n (\widehat{U}_i - \overline{\widehat{U}})(\widehat{V}_i - \overline{\widehat{V}})}{\sqrt{\sum_{i=1}^n (\widehat{V}_i - \overline{\widehat{V}})^2 \sum_{i=1}^n (\widehat{U}_i - \overline{\widehat{U}})^2}}, \quad \overline{\widehat{V}} = \frac{1}{n}\sum_{i=1}^n \widehat{V}_i, \quad \overline{\widehat{U}} = \frac{1}{n}\sum_{i=1}^n \widehat{U}_i;$$

or (b) the sample analog of (2.3), that is $\check{\rho}_C = 12 \sum_{i=1}^n (\widehat{U}_i - .5)(\widehat{V}_i - .5)/n$.

Comment 4.1 (Computation)**.** If the set $\mathcal{R}_n$ contains many elements, in step (2) we can either replace it by a smaller fine mesh or use a computationally fast method similar to Chernozhukov et al. (2022) to speed-up computation.[13] Note that the optimization program to obtain $\widehat{\alpha}_R(\bar{r})$ in step (3) only needs to be solved twice, once for r_0 in the upper tail and once for r_0 in the lower tail. Also, we require $m \geqslant 30$, which is thought to be the minimal sample size required to estimate one parameter.

4.3. **Bootstrap Inference.** Section 6 shows that the estimators described in Algorithm 1 follow normal distributions in large samples. The variances of these distributions, however, have complicated forms and are difficult to estimate. Section 6 also shows that the asymptotic distributions can be consistently estimated using exchangeable bootstrap. Exchangeable bootstrap is a general resampling method that includes empirical, weighted, wild and subsampling bootstrap as special cases; see Comment 4.2. The following algorithm describes how to obtain bootstrap draws of the estimators of ρ_C.

Algorithm 2 (Exchangeable Bootstrap Draws of Estimators)**.**

(1) Draw a realization of the weights $(\omega_{n1}, \dots, \omega_{nn})$ from a distribution that satisfies Assumption 6.2 in Section 6. Normalize the weights to add-up to one.

(2) Obtain a bootstrap draw of $\widehat{\beta}_R(r)$ at $r \in \bar{\mathcal{R}}_n$ by weighted DR, that is,

$$\widehat{\beta}_R^*(r) \in \operatorname{argmax}_{b \in \mathbb{R}^{d_x}} \sum_{i=1}^{n} \omega_{ni} \left[1(R_i \leqslant r) \log \Lambda(X_i'b) + 1(R_i > r) \log \Lambda(-X_i'b) \right].$$

[13]By stochastic equicontinuity of the conditional distribution processes $(r, x) \mapsto \sqrt{n}\left[\Lambda(x'\widehat{\beta}_R(r)) - \Lambda(x'\beta_R(r))\right]$, $R \in \{Y, W\}$, in Lemma 6.1, the meshwidth δ should be such that $\delta\sqrt{n} \to 0$ as $n \to \infty$.

(3) Obtain a bootstrap draw of $\widehat{\beta}_R(r)$ at $r \in \mathcal{R}_n \setminus \bar{\mathcal{R}}_n$ by restricted weighted DR, that is,

$$\widehat{\beta}^*_R(r) = (r - \bar{r})\widehat{\alpha}^*_R(\bar{r}) + x'\widehat{\beta}^*_R(\bar{r}),$$

where $\bar{r} := \arg\min_{r' \in \bar{\mathcal{R}}_n} |r - r'|$,

$$\widehat{\alpha}^*_R(\bar{r}) \in \operatorname{argmax}_{a \in \mathbb{R}} \sum_{i=1}^n \omega_{ni} \Big[1(R_i \leqslant r_0) \log \Lambda((r_0 - \bar{r})a + X_i'\widehat{\beta}^*_R(\bar{r})) \\ +1(R_i > r_0) \log \Lambda(-(r_0 - \bar{r})a - X_i'\widehat{\beta}^*_R(\bar{r})) \Big],$$

and $r_0 \in \mathcal{R}_n \setminus \bar{\mathcal{R}}_n$ is the same as in Algorithm 1.

(4) Obtain bootstrap draws of the estimators of the conditional ranks

$$\widehat{U}^*_i = \Lambda(X_i'\widehat{\beta}^*_Y(Y_i)), \quad \widehat{V}^*_i = \Lambda(X_i'\widehat{\beta}^*_W(W_i)).$$

(5) Obtain a bootstrap draw of the estimator of ρ_C as either (a) the weighted sample correlation, that is

$$\widehat{\rho}^*_C = \frac{\sum_{i=1}^n \omega_{ni}(\widehat{U}^*_i - \overline{\widehat{U}}^*)(\widehat{V}^*_i - \overline{\widehat{V}}^*)}{\sqrt{\sum_{i=1}^n \omega_{ni}(\widehat{V}^*_i - \overline{\widehat{V}}^*)^2 \sum_{i=1}^n \omega_{ni}(\widehat{U}^*_i - \overline{\widehat{U}}^*)^2}},$$

where $\overline{\widehat{V}}^* = \sum_{i=1}^n \omega_{ni}\widehat{V}^*_i/n$ and $\overline{\widehat{U}}^* = \sum_{i=1}^n \omega_{ni}\widehat{U}^*_i/n$; or (b) the weighed sample analog of (2.3),

$$\breve{\rho}^*_C = \frac{12}{n}\sum_{i=1}^n \omega_{ni}(\widehat{U}^*_i - .5)(\widehat{V}^*_i - .5).$$

Comment 4.2 (Bootstrap Weights)**.** van der Vaart et al. (1996) notes that by appropriately selecting the distribution of the weights, exchangeable bootstrap covers the most common bootstrap schemes as special cases. The empirical bootstrap corresponds to where $(w_{n1}, ..., w_{nn})$ is a multinomial vector with parameter n and probabilities $(1/n, ..., 1/n)$. The weighted bootstrap corresponds to where $w_{n1}, ..., w_{nn}$ are i.i.d. nonnegative random variables with $\mathrm{E}(w_{n1}) = \mathrm{Var}(w_{n1}) = 1$, e.g. standard exponential. The wild bootstrap corresponds to where $w_{n1}, ..., w_{nn}$ are i.i.d. vectors with $\mathrm{E}(w_{n1}^{2+\varepsilon}) < \infty$ for some $\varepsilon > 0$, and $\mathrm{Var}(w_{n1}) = 1$. The m out of n bootstrap corresponds to letting $(w_{n1}, ..., w_{nn})$ be equal to $\sqrt{n/m}$ times multinomial vectors with parameter m and probabilities $(1/n, ..., 1/n)$. The subsampling bootstrap corresponds to letting $(w_{n1}, ..., w_{nn})$ be a row in which the number $n(n-m)^{-1/2}m^{-1/2}$ appears m times and 0 appears $n - m$ times ordered at random, independent of the data.

We now show how to use the exchangeable bootstrap to obtain standard errors for the estimators of ρ_C and construct asymptotic confidence intervals for ρ_C. Algorithm 3 describes the procedure for $\widehat{\rho}_C$. A similar algorithm applies to $\breve{\rho}_C$. Let B a prespecified number of bootstrap repetitions and α be the significance level for the confidence intervals. For example, $B = 500$ and $\alpha = 0.05$.

Algorithm 3 (Inference on ρ_C based on $\widehat{\rho}_C$)**.**

(1) Draw $\{\widehat{Z}^*_b : 1 \leqslant b \leqslant B\}$ as i.i.d. realizations of $\widehat{Z}^* = \sqrt{n}\,(\widehat{\rho}^*_C - \widehat{\rho}_C)$ using Algorithms 1 and 2.

(2) Compute a bootstrap estimate of the asymptotic standard deviation of $\widehat{\rho}_C$, σ_ρ, such as the bootstrap interquartile range rescaled by the normal distribution:

$$\widehat{\sigma}_\rho = \frac{q_{.75} - q_{.25}}{z_{.75} - z_{.25}},$$

where q_p is the p-th quantile of $\{\widehat{Z}^*_{\rho,b} : 1 \leqslant b \leqslant B\}$ and z_p is the p-th quantile of $N(0,1)$.

(3) Compute B bootstrap draws of the T-statistic, $\{T_b : 1 \leqslant b \leqslant B\}$, where $T_b = |\widehat{Z}^*_{\rho,b}|/\widehat{\sigma}_\rho$.

(4) Construct an asymptotic $(1-\alpha)$-confidence interval for ρ_C as:

$$\mathrm{ACI}_{1-\alpha}(\rho_C) = \widehat{\rho}_C \pm \widehat{t}_{1-\alpha}\widehat{\sigma}_\rho/\sqrt{n},$$

where $\widehat{t}_{1-\alpha}$ is the $(1-\alpha)$-quantile of $\{T_b : 1 \leqslant b \leqslant B\}$.

5. Empirical Application

We analyze intergenerational income mobility in Switzerland using the Economic Well-Being of the Working and Retirement Age Population Data (WiSiER).

5.1. **Data.** WiSiER data include Swiss individuals from 11 Cantons from 1982 to 2016. The Swiss Federal Statistical Office merged data from tax records, social insurance, unemployment data, and surveys, creating a unique opportunity to analyze mobility. An ID can match parents and children. While many approaches seem feasible, we compare fathers and children at the same age of 35. As a result, the observations stem from different periods, with most of our successful matches coming from 1982-1990 (fathers) and 2000-2016 (children). The primary outcome variable is yearly real insured labor income (AHV) in $1,000$ Swiss francs (CHF) at the age of 35. The following covariates are available for both fathers and children: months experience, indicators for high-education (12 or more years of schooling), Swiss citizenship, and being single, and number of own children. Further, we include the fathers age at child's birth, and year and canton fixed effects for the children. Finally, for the analysis we exclude the following observations: (i) children where there is no parent in the data, (ii) observations with no information on the child's or father's birth year, and (iii) whenever the father was younger than 15 at the birth of the child. We conduct separate analyses for the relationships with sons and daughters. Table 4 reports descriptive statistics for the data used in the analysis. It shows that father's characteristics are similar in families with sons and daughters. This alleviates a potential concern about endogenous selection in the comparison between sons and daughters.

5.2. **Rank-Rank Regressions.** Table 5 reports the results of RRR and CRRR. The CRRR results are obtained using Algorithms 1 and 3 for the correlation-based estimator with a logistic link function and a mesh of 200 points located at sample quantiles in a sequence of orders from 0.01 to 0.99 with increments of $0.98/199$. We use linear interpolation to obtain estimates of the conditional ranks

Table 4. Descriptive Statistics

	Father-Son				Father-Daughter			
	Son		Father		Daughter		Father	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Income (1,000 CHF)	91	43	80	42	52	35	80	42
Age at birth			26.8	3.3			26.9	3.3
Higher Education	0.56	0.50	0.34	0.47	0.50	0.50	0.34	0.48
Months of Experience	191	30	50	34	184	31	51	34
Swiss Citizen	0.96	0.20	0.88	0.33	0.96	0.21	0.87	0.34
Single	0.46	0.50	0.16	0.37	0.43	0.49	0.17	0.38
Number of Children	1.20	1.11	2.45	0.86	1.35	1.10	2.46	0.87

Notes: sample size is $10,363$ *for father-son and* $9,581$ *for father-daughter.*

corresponding to intermediate points in the mesh. The standard errors (SE) and 95% confidence intervals (95% CI) are computed by empirical bootstrap with 500 repetitions. Based on the results of numerical simulations reported in Appendix C, we do not impose tail restrictions. In results not reported, we find very similar estimates, standard errors and confidence intervals for regression-based and fully restricted estimators.[14] We show the robustness of the results to the choice of link function in Section 5.5.

We find significant positive income persistence in both father-son and father-daughter relationships, with and without covariates. However, the persistence is much stronger for sons than for daughters suggesting the presence of a gender gap in intergenerational transmission of income even after controlling for the father's and child's characteristics. Comparing RRR and CRRR, we find that within-group persistence accounts for approximately 62% of the overall income persistence for sons and about 52% for daughters. These results highlight the substantial role of both within-group and between-group differences in explaining intergenerational mobility.

A subgroup analysis reveals relatively more mobility in families with a larger number of children and with a low educated father. In particular, we find relatively less persistence for sons in large families and more for daughters of high educated fathers. This would be consistent with decreasing returns of intergenerational transfers with respect to family size and increasing with respect to father's education. This heterogeneity, however, is not statistically significant. We do not find differences in intergenerational mobility for families with immigrant fathers in Switzerland, unlike the results of Abramitzky et al. (2021) for the U.S. This difference might be due to the small fraction of immigrant fathers in the sample, see Table 4.

[14]These results are available from the authors upon request.

TABLE 5. Intergenerational Income Mobility in Switzerland

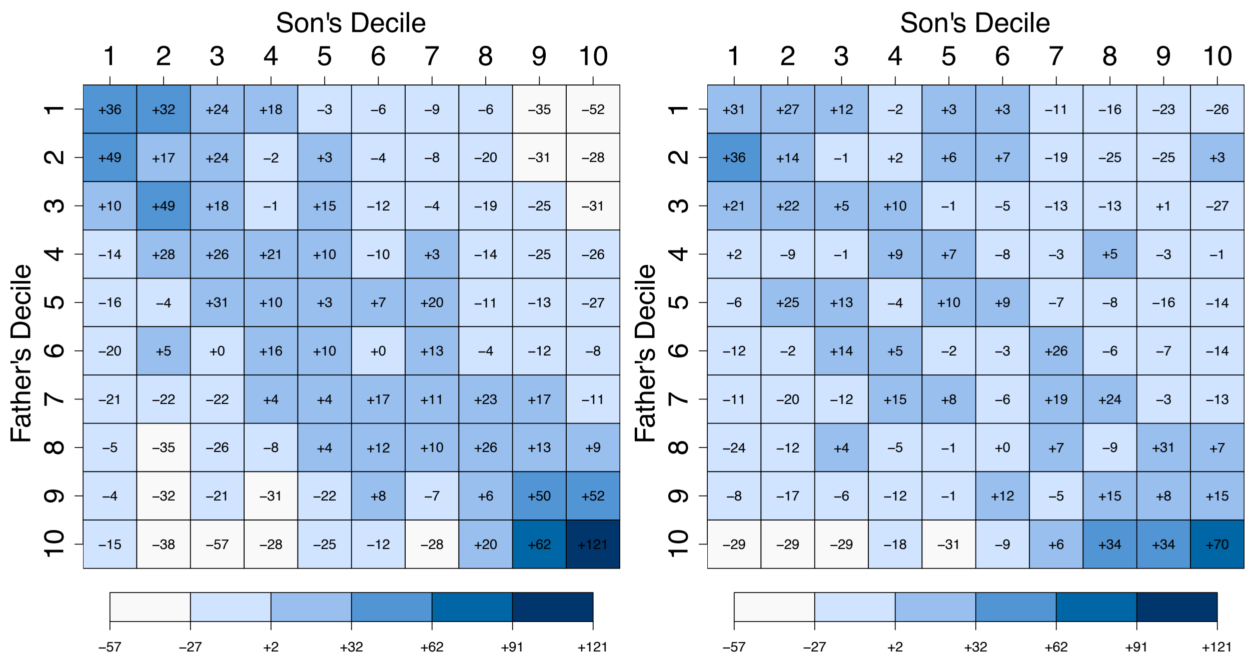
	Father-Son				Father-Daughter			
	Coef.	SE	95% CI		Coef.	SE	95% CI	
RRR	0.202	0.010	0.182	0.222	0.088	0.010	0.069	0.107
CRRR	0.126	0.011	0.106	0.147	0.046	0.010	0.027	0.066
CRRR, by Father:								
High education	0.131	0.017	0.095	0.167	0.085	0.016	0.014	0.156
Age >26 at birth	0.130	0.015	0.100	0.160	0.054	0.014	0.021	0.087
Swiss citizen	0.128	0.012	0.105	0.152	0.045	0.011	0.024	0.065
More >2 children	0.109	0.016	0.070	0.149	0.044	0.016	0.012	0.077

Notes: Correlation-based estimator with logistic link function and a mesh of 200 points. SE and 95% CI obtained by empirical bootstrap with 500 repetitions. Covariates include father's and child's months of experience, higher education, Swiss citizenship, single and number of own children; father's age birth; and child's year and canton fixed effects. Sample size is $10,363$ *for father-son and* $9,581$ *for father-daughter data.*

5.3. **Transition Matrices.** Figures 1 and 2 show heatmaps of transition matrices for father-son and father-daughter, respectively. These matrices are a parsimonious representation of the joint distribution of income for father and child discretized in cells defined by deciles. They are commonly used in intergenerational mobility studies to provide a more granular measure of persistence than the rank-rank regressions. We report all the entries in percent deviations from 0.1 because all the entries should be equal to 0.1 under perfect mobility, that is when the income of the child is independent of the income of the father. Panels (A) report transition matrices based on marginal ranks, similar to previous studies. Panels (B) report conditional transition matrices based on conditional ranks, which are new to this paper and capture within-group dependence. For father-son, we find that the highest values in panel (A) are concentrated on the diagonal, which is consistent with the positive RRR estimate in Table 5. The results in panel (B) show a less clear pattern once we control for covariates, consistent with the lower CRRR estimates in Table 5. The results for father-daughter show similar but weaker patterns as we expect from the smaller correlation estimates in Table 5. Interestingly, for both sons and daughters the highest probability occurs at the bottom right corner of the very top deciles conditionally and unconditionally.

5.4. **Rank-Rank Regressions Excluding Child's Covariates.** One concern about the CRRR results in Table 5 is that the child's covariates might be picking up indirect sources of intergenerational mobility of income. For example, fathers might invest in child's education to increase the child's income prospects. To deal with this concern, Table 6 reports CRRR results where the child's covariates, other than year and canton fixed effects, are excluded from the covariate set X. These

Figure 1. Transition matrices: Father-Son



(a) Marginal ranks (b) Conditional ranks

Notes: entries are in percent deviations from 0.01. *Sample size is* 10,363 *for father-son and* 9,581 *for father-daughter.*

results are obtained using the correlation-based estimator with a logistic link function with the same parameter choices as in Table 5.

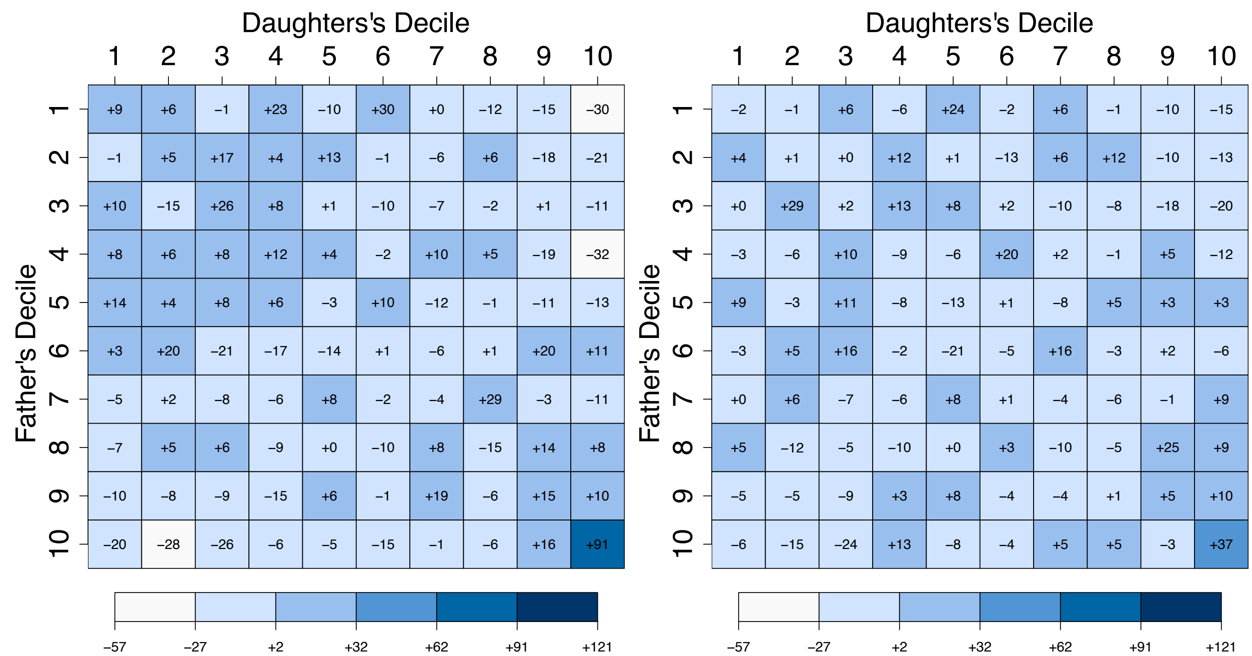
As expected, not accounting for the child's covariates increases the importance of within-group persistence with the estimates increasing to about 80% for father-son and 69% for father-daughter. In both cases the increase is about 17-18%. The other conclusions remain unchanged. In particular, we still find a significant gender gap in intergenerational transmission of income, and relatively less persistence for sons in large families and more for daughters of high educated fathers.

5.5. **Robustness to Link Function.** Table 7 reports the results of CRRR using the correlation-based estimator with a Gaussian or probit link function. The estimates, standard errors and confidence intervals are almost identical to Table 5 showing the robustness of the results to the use of the logistic versus Gaussian link functions.

6. Asymptotic Theory

In this section we provide asymptotic theory for the estimators of the CRRR slope ρ_C. We focus on the correlation-based and fully-restricted estimators of Algorithm 1. We derive their asymptotic

FIGURE 2. Transition matrices: Father-Daughter



(A) Marginal ranks (B) Conditional ranks

Notes: entries are in percent deviations from 0.01. *Sample size is* 10, 363 *for father-son and* 9, 581 *for father-daughter.*

TABLE 6. Estimates Excluding Child's Covariates

	Father-Son				Father-Daughter			
	Coef.	SE	95% CI		Coef.	SE	95% CI	
RRR	0.202	0.010	0.182	0.222	0.088	0.010	0.069	0.107
CRRR	0.161	0.010	0.140	0.182	0.061	0.010	0.041	0.081
CRRR, by Father:								
High education	0.166	0.016	0.132	0.200	0.089	0.015	0.031	0.147
Age >26 at birth	0.164	0.014	0.135	0.194	0.077	0.013	0.038	0.116
Swiss citizen	0.164	0.012	0.140	0.187	0.060	0.010	0.039	0.081
More >2 children	0.142	0.017	0.100	0.183	0.064	0.016	0.031	0.098

Notes: Correlation-based estimator with logistic link function and a mesh of 200 points. SE and 95% CI obtained by empirical bootstrap with 500 repetitions. Covariates include father's months of experience, higher education, Swiss citizenship, single and number of own children; father's age birth; and child's year and canton fixed effects. Sample size is 10, 363 *for father-son and* 9, 581 *for father-daughter data.*

Table 7. Robustness to Link Function: Probit Estimates

	Father-Son				Father-Daughter			
	Coef.	SE	95% CI		Coef.	SE	95% CI	
RRR	0.202	0.010	0.182	0.222	0.088	0.010	0.069	0.107
CRRR	0.127	0.011	0.107	0.148	0.047	0.010	0.028	0.067
CRRR, by Father:								
High education	0.132	0.017	0.096	0.168	0.086	0.016	0.015	0.157
Age >26 at birth	0.131	0.014	0.102	0.161	0.055	0.015	0.023	0.087
Swiss citizen	0.129	0.012	0.106	0.153	0.045	0.011	0.025	0.066
More >2 children	0.110	0.016	0.071	0.150	0.046	0.016	0.014	0.079

Notes: Correlation-based estimator with Gaussian link function and a mesh of 200 points. SE and 95% CI obtained by empirical bootstrap with 500 repetitions. Covariates include father's and child's months of experience, higher education, Swiss citizenship, single and number of own children; father's age birth; and child's year and canton fixed effects. Sample size is $10,363$ *for father-son and* $9,581$ *for father-daughter data.*

distributions by the delta method. For example, we take the following steps for the correlation-based estimator:

(1) Express the parameter ρ_C as a correlation-based functional of the function-valued inputs $F_{Y|X}$, $F_{W|X}$ and F_Z, where $Z = (Y, W, X')'$. That is:

$$\rho_C = \phi(F_{Y|X}, F_{Y|X}, F_Z) := \frac{\int [F_{Y|X}(y \mid x) - .5][F_{W|X}(w \mid x) - .5] \mathrm{d}F_Z(z)}{\sqrt{\int [F_{W|X}(w \mid x) - .5]^2 \mathrm{d}F_Z(z) \int [F_{Y|X}(y \mid x) - .5]^2 \mathrm{d}F_Z(z)}}. \tag{6.1}$$

(2) Show that the plug-in estimator of ρ_C using ϕ, $\tilde{\rho}_C$, is a restricted correlation-based estimator:

$$\tilde{\rho}_C = \phi(\widehat{F}_{Y|X}, \widehat{F}_{W|X}, \widehat{F}_Z) = \frac{\sum_{i=1}^n (\widehat{U}_i - .5)(\widehat{V}_i - .5)}{\sqrt{\sum_{i=1}^n (\widehat{V}_i - .5)^2 \sum_{i=1}^n (\widehat{U}_i - .5)^2}},$$

where $\widehat{F}_{Y|X}$ and $\widehat{F}_{W|X}$ are the DR estimators of $F_{Y|X}$, $F_{W|X}$ and $\widehat{F}_Z$ is the empirical distribution function of Z.

(3) Establish that the map ϕ is Hadamard differentiable in the relevant functional spaces at $(F_{Y|X}, F_{Y|X}, F_Z)$ with the affine and continuous derivative operator:

$$(z_Y, z_W, g_Z) \mapsto \phi'_{F_{Y|X}, F_{Y|X}, F_Z}(z_Y, z_W, g_Z),$$

where z_Y, z_W and g_Z are the limits of converging deviations from $F_{Y|X}$, $F_{Y|X}$ and F_Z.

(4) Apply the functional delta method to obtain the limit of $\tilde{\rho}_C$ from the limit of the deviations of the estimators of the inputs $\sqrt{n}(\widehat{F}_{Y|X} - F_{Y|X})$, $\sqrt{n}(\widehat{F}_{W|X} - F_{W|X})$, and $\sqrt{n}(\widehat{F}_Z - F_Z)$,

$$\sqrt{n}(\tilde{\rho}_C - \rho_C) \rightsquigarrow \phi'_{F_{Y|X}, F_{W|X}, F_Z}(Z_Y, Z_W, G_Z),$$

where $\rightsquigarrow$ and (Z_Y, Z_W, G_Z) are defined below.

(5) Show that $\widehat{\rho}_C$ has the same asymptotic distribution as $\tilde{\rho}_C$,

$$\sqrt{n}(\widehat{\rho}_C - \tilde{\rho}_C) \to_P 0.$$

The distribution of the fully-restricted estimator is derived following steps (1)-(4), and replacing the correlation-based functional in step (1) by the fully-restricted functional:

$$\rho_C = 12\, \phi_1(F_{Y|X}, F_{Y|X}, F_Z) := 12 \int [F_{Y|X}(y \mid x) - .5][F_{W|X}(w \mid x) - .5]\mathrm{d}F_Z(z). \tag{6.2}$$

Comment 6.1 (Regression-Based Estimators)**.** The limit distribution of the regression-based estimators is derived following analogous steps to the correlation-based estimator, and replacing the correlation-based representation of the functional in step (1) by the regression-based representation:

$$\rho_C = \varphi(F_{Y|X}, F_{Y|X}, F_Z) := \frac{\int [F_{Y|X}(y \mid x) - .5][F_{W|X}(w \mid x) - .5]\mathrm{d}F_Z(z)}{\int [F_{W|X}(w \mid x) - .5]^2 \mathrm{d}F_Z(z)}. \tag{6.3}$$

We provide the corresponding results in Appendix A.

Before stating formally the main results, we review the existing theory for the estimator of the RRR slope. The purpose of this review is to explain why the existing results do not cover the estimators of the CRRR slope. Hoeffding (1948) first derived the asymptotic distribution of the RRR slope estimator using the theory of U-statistics. We cannot follow the same approach because none of our estimators has a U-statistic representation. Ren and Sen (1995) alternatively derived the asymptotic distribution of the RRR slope estimator using the delta method. Ren and Sen (1995) used analogous steps to our procedure described above. The following remarks explain each step of our procedure and point out the challenges and differences with respect to Ren and Sen (1995).

Comment 6.2 (Functional Representation of ρ_C)**.** The functional and the inputs of the functional representation of the RRR slope, ρ, are different from ρ_C. In particular, Ren and Sen (1995) showed that:

$$\rho = \tilde{\phi}(F_{Y,W}) = 12 \int [F_{Y,W}(y, +\infty) - .5][F_{Y,W}(+\infty, w) - .5]\mathrm{d}F_{Y,W}(y, w).$$

The functional $\tilde{\phi}$ is an special case of the fully-restricted functional ϕ_1 in (6.2) where there are no covariates X. In the case of RRR, depite being a regression-based estimator, the denominator simplifies because the sample variances of the estimated marginal ranks are deterministic when Y and W are continuous.[15] This simplification is not available for the correlation-based and regression-based estimators of ρ_C because the sample variances of the estimated conditional ranks are random.

Comment 6.3 (Plug-in Estimator)**.** The plug-in estimator of ρ_C using ϕ is:

$$\phi(\widehat{F}_{Y|X}, \widehat{F}_{W|X}, \widehat{F}_Z) = \frac{\int [\Lambda(x'\widehat{\beta}_Y(y)) - .5][\Lambda(x'\widehat{\beta}_W(w)) - .5]\mathrm{d}\widehat{F}_Z(z)}{\int [\Lambda(x'\widehat{\beta}_W(w)) - .5]^2 \mathrm{d}\widehat{F}_Z(z)} = \tilde{\rho}_C,$$

[15] Indeed, these sample variances are equal to $(n^2-1)/(12n^2)$, see Ren and Sen (1995); and the regression-based and correlation-based versions of the RRR estimator are numerically identical if there are no ties in the observations of W and Y.

where the second equality follows from the properties of the empirical distribution function $\widehat{F}_Z$. This proves Step (2) above.

Comment 6.4 (Hadamard Differentiability of ϕ)**.** The argument to establish differentiability of the RRR functional $\tilde{\phi}$ does not apply to the CRRR functional ϕ for several reasons. First, the expression of ϕ is different from $\tilde{\phi}$. Second, the inputs and their estimators are also different. Moreover, the estimators of the inputs of the CRRR functional live in more complicated spaces than the estimators of the inputs of the RRR functional. Thus, while the estimator of F_Z lives in the space of Cadlag functions, the estimators of $F_{Y|X}$ and $F_{W|X}$ live in the space of bounded functions, but have limits in the space of continuous functions, once properly recentered and rescaled. Because of this difference, we need to establish Hadamard differentiability in the space of bounded functions, tangentially to the space of continuos functions.

Comment 6.5 (Limit Process of Input Estimators)**.** To apply the delta method, we need to characterize the limit process for the estimator of the inputs. This characterization is much more challenging for CRRR than RRR. Thus, for example, Ren and Sen (1995) can rely on existing functional central limit theorems for the empirical distribution to establish the limit process over the entire support of Y and W. Unfortunately, the existing functional central limit theorems for DR estimators of conditional distributions have only been established on a compact strict subset of the support of Y and W; see, for example, Chernozhukov et al. (2013). We deal with this challenge by imposing restrictions on the DR model of the conditional distributions at the tails. These restrictions allow us to extend the central limit theorems to the entire support of Y and W. In numerical simulations, however, we find that estimators with and without imposing the tail restrictions perform similarly in terms of bias, standard deviation and root mean squared error.

We formally state now the main results from the steps (4) and (5). The result from step (3) is relegated to Appendix B.1 because it is of more technical nature. We state all the results for the logistic link function because it produces analytically simpler expressions, but it can be readily extended to the Gaussian link at the cost of more cumbersome notation.

We start by imposing some conditions on the DR model.

Assumption 6.1 (DR Model)**.** *For $R \in \{Y, W\}$: (a) The conditional distribution function takes the form $F_{R|X}(r \mid x) = \Lambda(x'\beta_R(r))$ for all $r \in \mathcal{R}$ and $x \in \mathcal{X}$, where $\Lambda(u) = (1 + \exp(-u))^{-1}$, the standard logistic distribution. (b) The support $\mathcal{R}$ is an open interval in $\mathbb{R}$ and the conditional density function $f_{R|X}(r \mid x)$ exists and is positive in (r, x) on (R, X); it is uniformly bounded and uniformly continuous in (r, x) on (R, X). (c) $E\|X\|^2 < \infty$ and the minimum eigenvalue of:*

$$J_R(r) := \mathrm{E}\left[\lambda(X'\beta_R(r))XX'\right],$$

is bounded away from zero uniformly over $r \in \mathcal{R}$, where $\lambda = \Lambda(1 - \Lambda)$ is the derivative of Λ. (d) Let $\beta_R(y)$ be partitioned as $(\beta_{R,1}(r), \beta_{R,-1}(r)')'$ where $\beta_{R,1}(r)$ is the intercept and $\beta_{R,-1}(r)$ includes the rest *of the components. Then, for* $r \in \mathcal{R} \setminus \bar{\mathcal{R}}$*, where* $\bar{\mathcal{R}}$ *is a closed subinterval of the interior of* $\mathcal{R}$*,* $\beta_{R,1}(r) = \beta_{R,1}(\bar{r}) + (r - \bar{r})\alpha_R(\bar{r})$ *for* $\bar{r} := \arg\min_{r' \in \bar{\mathcal{R}}} |r - r'|$ *and some* $\alpha_R(\bar{r}) > 0$*, and* $\beta_{R,-1}(r) = \beta_{R,-1}(\bar{r})$*.*

Comment 6.6 (DR Model)**.** The conditions in Assumption 6.1(a)-(c) are the same as in Chernozhukov et al. (2013). They are used to obtain a functional central limit for the DR estimator of the conditional distribution $F_{R|X}(r \mid x)$ on $\bar{\mathcal{R}}$. Assumption 6.1(d) imposes restrictions on the tails that allow us to extend the functional central limit theorem to $\mathcal{R}$.

In order to state the result about the limit process for the inputs, we define, for $R \in \{Y, W\}$,

$$\begin{aligned}\ell_{r,x}(R, X) &= 1\{r \in \bar{\mathcal{R}}\}\lambda(x'\beta_R(r))x' J_R^{-1}(r) \left[\Lambda(X'\beta_R(r)) - 1\{R \leqslant r\}\right] X \\ &\quad + 1\{r \in \mathcal{R} \setminus \bar{\mathcal{R}}\}\lambda(x'\beta_R(r)) \left\{ \frac{r - \bar{r}}{r_0 - \bar{r}} \frac{\Lambda(X'\beta_R(r_0)) - 1\{R \leqslant r_0\}}{\mathrm{E}[\lambda(X'\beta_R(r_0))]} \right. \\ &\quad \left. + \left[x - \frac{r - \bar{r}}{r_0 - \bar{r}} \frac{\mathrm{E}[\lambda(X'\beta_R(r_0))X]}{\mathrm{E}[\lambda(X'\beta_R(r_0))]} \right]' J_R^{-1}(\bar{r}) \left[\Lambda(X'\beta_R(\bar{r})) - 1\{R \leqslant \bar{r}\}\right] X \right\}.\end{aligned}$$

Consider the empirical processes $(r, x) \mapsto \widehat{Z}_R(r, x) := \sqrt{n}\left(\widehat{F}_{R|X}(r \mid x) - F_{R|X}(r \mid x)\right)$, $R \in \{Y, W\}$, and $f \mapsto \widehat{G}_Z(f) := \sqrt{n} \int f \mathrm{d}(\widehat{F}_Z - F_Z)$, where $\widehat{F}_{R|X}(r \mid x) := \Lambda(x'\widehat{\beta}_R(r))$, $\widehat{F}_Z$ is the empirical distribution function of $Z = (Y, W, X)$, and $\mathcal{F}$ is a class of measurable functions that (i) includes $F_{Y|X}$, $F_{W|X}$, $F_{Y|X}^2$, $F_{W|X}^2$, $F_{Y|X}F_{W|X}$ and the indicators of all the rectangles in $\bar{\mathbb{R}}^{d_x+2}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is the extended real line, and (ii) is totally bounded under the metric:

$$\lambda(f, \tilde{f}) = \left[\int (f - \tilde{f})^2 \mathrm{d}F_Z\right]^{1/2}, \quad f, \tilde{f} \in \mathcal{F}.$$

Let $Z_n \rightsquigarrow Z$ in $\mathbb{E}$ denote weak convergence of a stochastic process Z_n to a random element Z in a normed space $\mathbb{E}$, as defined in van der Vaart et al. (1996).

Lemma 6.1 (Limit Processses for Inputs)**.** *Assume that Assumption 6.1 holds, the support* $\mathcal{X}$ *is a compact subset of* $\mathbb{R}^{d_x}$*, and* $\{Z_i = (Y_i, W_i, X_i)\}_{i=1}^n$ *is a random sample of* $Z = (Y, W, X)$*. Then, in the metric space* $\ell^\infty(\mathcal{YWXF})$*,*

$$(\widehat{Z}_Y(y, x), \widehat{Z}_W(w, x), \widehat{G}_Z(f)) \rightsquigarrow (Z_Y(y, x), Z_W(w, x), G_Z(f))$$

as stochastic processes indexed by (y, w, x, f)*. The limit process is a zero-mean tight Gaussian process such that:*

$$Z_R(r, x) = \mathbb{G}(\ell_{r,x}), \quad R \in \{Y, W\}, \text{ and } G_Z(f) = \mathbb{G}(f),$$

where $\mathbb{G}$ *is a* P*-Brownian bridge.*

The next result states a central limit theorem for $\tilde{\rho}_C$ and $\breve{\rho}_C$, and the asymptotic equivalence between $\tilde{\rho}_C$ and $\widehat{\rho}_C$.

Theorem 6.1 (Limit Distribution of $\widehat{\rho}_C$, $\tilde{\rho}_C$ and $\breve{\rho}_C$)**.** *Under the conditions of Lemma 6.1: (1) in* $\mathbb{R}$*,*

$$\sqrt{n}\,(\tilde{\rho}_C - \rho_C) \rightsquigarrow Z_\rho := 12\,[Z_{1,\rho} - \rho_C(Z_{2,\rho} + Z_{3,\rho})/2] \quad \text{and} \quad \sqrt{n}\,(\breve{\rho}_C - \rho_C) \rightsquigarrow 12 Z_{1,\rho},$$

where $Z_{1,\rho}$, $Z_{2,\rho}$ and $Z_{3,\rho}$ are zero-mean Gaussian random variables defined by

$$Z_{1,\rho} := \int \left\{ Z_Y(y,x)[F_{W|X}(w \mid x) - .5] + Z_W(w,x)[F_{Y|X}(y \mid x) - .5] \right\} \mathrm{d}F_Z(y,w,z) + G_Z\left([F_{Y|X} - .5][F_{W|X} - .5]\right),$$

$$Z_{2,\rho} := 2\int Z_W(w,x)[F_{W|X}(w \mid x) - .5]\mathrm{d}F_Z(y,w,z) + G_Z\left([F_{W|X} - .5]^2\right),$$

and

$$Z_{3,\rho} := 2\int Z_Y(y,x)[F_{Y|X}(y \mid x) - .5]\mathrm{d}F_Z(y,w,z) + G_Z\left([F_{Y|X} - .5]^2\right).$$

(2) $\widehat{\rho}_C$ *has the same limit distribution as* $\tilde{\rho}_C$ *because*

$$\sqrt{n}\left(\widehat{\rho}_C - \tilde{\rho}_C\right) \to_{\mathrm{P}} 0.$$

The variance of the limit processes $Z_{1,\rho}$ and Z_ρ have complicated expressions that might be difficult to estimate analytically. To avoid this difficulty, we propose the use of bootstrap to make inference. We show that the exchangeable bootstrap draws of Algorithm 2 have the same asymptotic distribution as the CRRR estimators under the following assumption on the weights:

Assumption 6.2 (Exchangeable Bootstrap)**.** *For each n, $(\omega_{n1}, ..., \omega_{nn})$ is an exchangeable,*[16] *nonnegative random vector, which is independent of the data, such that for some $\epsilon > 0$*

$$\sup_n \mathrm{E}[\omega_{n1}^{2+\epsilon}] < \infty, \;\; n^{-1}\sum_{i=1}^{n}(\omega_{ni} - \bar{\omega}_n)^2 \to_{\mathrm{P}} 1, \;\; \bar{\omega}_n \to_{\mathrm{P}} 1, \tag{6.4}$$

where $\bar{\omega}_n = n^{-1}\sum_{i=1}^n \omega_{ni}$.

In order to state the results about bootstrap validity formally, we follow the notation and definitions in van der Vaart et al. (1996). Let D_n denote the data vector and M_n be the vector of random variables used to generate bootstrap draws given D_n. Consider the random element $\mathbb{Z}_n^* = \mathbb{Z}_n(D_n, M_n)$ in a normed space $\mathbb{E}$. We say that the bootstrap law of $\mathbb{Z}_n^*$ consistently estimates the law of some tight random element $\mathbb{Z}$ and write $\mathbb{Z}_n^* \rightsquigarrow_{\mathrm{P}} \mathbb{Z}$ in $\mathbb{E}$ if:

$$\sup_{h \in \mathrm{BL}_1(\mathbb{E})} \left|\mathrm{E}_{M_n} h\left(\mathbb{Z}_n^*\right) - \mathrm{E}h(\mathbb{Z})\right| \to_{\mathrm{P}} 0, \tag{6.5}$$

where $\mathrm{BL}_1(\mathbb{E})$ denotes the space of functions with Lipschitz norm at most 1 and E_{M_n} denotes the conditional expectation with respect to M_n given the data D_n; and $\to_{\mathrm{P}}$ denotes convergence in (outer) probability.

We now provide a bootstrap central limit theorem for the estimators of the CRRR slope. This result follows from a functional central limit theorem for the input processes, which we establish in Lemma B.4 in Appendix B, and the functional delta method for the bootstrap.

[16] A sequence of random variables $X_1, X_2, ..., X_n$ is exchangeable if for any finite permutation σ of the indices $1, 2, ..., n$ the joint distribution of the permuted sequence $X_{\sigma(1)}, X_{\sigma(2)}, ..., X_{\sigma(n)}$ is the same as the joint distribution of the original sequence.

Theorem 6.2 (Exchangeable Bootstrap Consistency)**.** *Under the conditions of Lemma 6.1 and Assumption 6.2: in* $\mathbb{R}$*,*

$$\sqrt{n}\,(\tilde{\rho}^*_C - \tilde{\rho}_C) \rightsquigarrow_{\mathrm{P}} Z_\rho, \quad \sqrt{n}\,(\widehat{\rho}^*_C - \widehat{\rho}_C) \rightsquigarrow_{\mathrm{P}} Z_\rho, \quad \textit{and} \quad \sqrt{n}\,(\breve{\rho}^*_C - \widehat{\rho}_C) \rightsquigarrow_{\mathrm{P}} Z_{1,\rho}$$

that is, exchangeable bootstrap consistently estimates the law of the limit processes Z_ρ *and* $Z_{1,\rho}$*. In particular,*

$$\widehat{\sigma}_\rho \to_{\mathrm{P}} \sigma_\rho \quad \textit{and} \quad \mathrm{P}\,\{\rho_C \in ACI_{1-\alpha}(\rho_C)\} \to 1 - \alpha \quad \textit{as} \quad n \to \infty,$$

where σ_ρ *is the standard deviation of the limit process* Z_ρ*, and* $\widehat{\sigma}_\rho$ *and* $ACI_{1-\alpha}(\rho_C)$ *are defined in Algorithm 3.*

7. CONCLUSION

This paper introduces the conditional rank-rank regression (CRRR) as an alternative to traditional rank-rank regressions with covariates (RRRX) for measuring within-group mobility and persistence. The CRRR uses conditional ranks of the variables of interest given covariates, in contrast to RRRX which uses marginal ranks net of covariate effects. We show that the CRRR slope preserves an intuitive interpretation as the average conditional rank correlation between the variables, similar to RRR without covariates. In contrast, the slope of RRRX loses the rank correlation interpretation and can take on values outside the interval $[-1, 1]$. The CRRR is also suitable for subgroup analysis, where the CRRR slopes maintain a rank correlation interpretation conditional on the groups.

We propose a distribution regression estimator for CRRR where the conditional distributions are modeled flexibly using parametric link functions. The estimator is easy to implement and computationally tractable. We derive asymptotic theory for the estimator based on the functional delta method. The analytic asymptotic variance is cumbersome, so we propose an exchangeable bootstrap procedure for inference. The bootstrap procedure is also used to construct confidence intervals. We illustrate the usefulness of CRRR in an empirical application to intergenerational income mobility in Switzerland. The application reveals stronger intergenerational persistence between fathers and sons than fathers and daughters, where the within-group persistence accounts for between 52% and 79% of the overall persistence. We also find some evidence of heterogeneity across groups defined by father's education and family size. The results are robust to the exclusion of child's covariates and the use of logistic or Gaussian link functions.

In summary, CRRR provides a well-grounded measure of within-group mobility and persistence. It also allows us to decompose the overall persistence captured by RRR into within-group persistence captured by CRRR plus a remainder term interpretable as between-group persistence. The distribution regression estimator, coupled with exchangeable bootstrap inference, provides a practical and flexible way to implement CRRR in empirical applications. We expect CRRR will be a useful addition to the toolkit of methods for studying mobility and persistence.

References

Abramitzky, R., L. Boustan, E. Jácome, and S. Pérez (2021): "Intergenerational mobility of immigrants in the United States over two centuries," *American Economic Review*, 111, 580–608.

Adermon, A., M. Lindahl, and D. Waldenström (2018): "Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations," *The Economic Journal*, 128, F482–F513.

Beller, E. and M. Hout (2006): "Intergenerational social mobility: The United States in comparative perspective," *The future of children*, 19–36.

Blomqvist, N. (1950): "On a measure of dependence between two random variables," *The Annals of Mathematical Statistics*, 593–600.

Chernozhukov, V. (2005): "Extremal quantile regression," *Ann. Statist.*, 33, 806–839.

Chernozhukov, V. and I. Fernández-Val (2011): "Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks," *Review of Economic Studies*, 78, 559–589.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2022): "Fast algorithms for the quantile regression process," *Empirical economics*, 1–27.

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, 1553–1623.

Chetverikov, D. and D. Wilhelm (2023): "Inference for rank-rank regressions," *arXiv preprint arXiv:2310.15512*.

Cramér, H. (1999): *Mathematical methods of statistics*, vol. 26, Princeton university press.

Dahl, M. W. and T. DeLeire (2008): *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*, University of Wisconsin-Madison, Institute for Research on Poverty Madison ….

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): "Modelling extremal events," 33.

Gijbels, I., N. Veraverbeke, and M. Omelka (2011): "Conditional copulas, association measures and their applications," *Computational Statistics & Data Analysis*, 55, 1919–1932.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *The Annals of Mathematical Statistics*, 19, 293–325.

Kendall, M. G. (1948): "Rank correlation methods." .

Kitagawa, T., M. Nybom, and J. Stuhler (2018): "Measurement error and rank correlations," Tech. rep., cemmap working paper.

Kruskal, W. H. (1958): "Ordinal measures of association," *Journal of the American Statistical Association*, 53, 814–861.

Lei, L. (2024): "Causal Interpretation of Regressions With Ranks," *arXiv preprint arXiv:2406.05548*.

Li, C. and B. E. Shepherd (2012): "A new residual for ordinal outcomes," *Biometrika*, 99, 473–480.

Liu, Q., C. Li, V. Wanga, and B. E. Shepherd (2018): "Covariate-adjusted Spearman's rank correlation with probability-scale residuals," *Biometrics*, 74, 595–605.

Maasoumi, E., L. Wang, and D. Zhang (2022): "Generalized Intergenerational Mobility Regressions," Tech. rep., Working paper, Emory University.

Murphy, R. and F. Weinhardt (2020): "Top of the class: The importance of ordinal rank," *The Review of Economic Studies*, 87, 2777–2826.

Ren, J.-J. and P. K. Sen (1995): "Hadamard differentiability on D [0, 1] p," *Journal of Multivariate Analysis*, 55, 14–28.

Shepherd, B. E., C. Li, and Q. Liu (2016): "Probability-scale residuals for continuous, discrete, and censored data," *Canadian Journal of Statistics*, 44, 463–479.

Spearman, C. (1904): "The Proof and Measurement of Association between Two Things," *The American Journal of Psychology*, 15, 72–101.

van der Vaart, A. W., J. A. Wellner, A. W. van der Vaart, and J. A. Wellner (1996): *Weak convergence*, Springer.

Veraverbeke, N., M. Omelka, and I. Gijbels (2011): "Estimation of a conditional copula and association measures," *Scandinavian Journal of Statistics*, 38, 766–780.

Appendix A. Regression-Based Estimators

Algorithm 4 (Regression-based Estimators)**.** Steps (1)–(3) are the same as in Algorithm 1. In step (4) estimate ρ_C as either (a) the slope of the linear regression of $\widehat{U}_i$ on $\widehat{V}_i$, that is

$$\widehat{\varrho}_C = \frac{\sum_{i=1}^n \widehat{U}_i(\widehat{V}_i - \overline{\widehat{V}})}{\sum_{i=1}^n (\widehat{V}_i - \overline{\widehat{V}})^2}, \quad \overline{\widehat{V}} = \frac{1}{n}\sum_{i=1}^n \widehat{V}_i;$$

or (b) the slope of the restricted linear regression of $\widehat{U}_i$ on $\widehat{V}_i$, that is

$$\tilde{\varrho}_C = \frac{\sum_{i=1}^n (\widehat{U}_i - .5)(\widehat{V}_i - .5)}{\sum_{i=1}^n (\widehat{V}_i - .5)^2}.$$

Theorem A.1 (Limit Distribution of $\tilde{\varrho}_C$ and $\widehat{\varrho}_C$)**.** *Under the conditions of Lemma 6.1: (1) in* $\mathbb{R}$*,*

$$\sqrt{n}\,(\tilde{\varrho}_C - \rho_C) \rightsquigarrow 12\,[Z_{1,\rho} - \rho_C Z_{2,\rho}],$$

where $Z_{1,\rho}$ and $Z_{2,\rho}$ are zero-mean Gaussian random variables defined in Theorem 6.1. (2) $\widehat{\varrho}_C$ has the same limit distribution as $\tilde{\varrho}_C$ because

$$\sqrt{n}\,(\widehat{\varrho}_C - \tilde{\varrho}_C) \to_{\mathrm{P}} 0.$$

Comment A.1 (Reverse Regression-Based Estimators)**.** The limit distribution of the reverse regression-based estimator can be trivially obtained from the regression-based case by relabeling the variables Y and W. Thus, let $\tilde{r}_C$ denote the reverse regression-based restricted estimator, that is

$$\tilde{r}_C = \frac{\sum_{i=1}^n (\widehat{U}_i - .5)(\widehat{V}_i - .5)}{\sum_{i=1}^n (\widehat{U}_i - .5)^2}.$$

By Theorem A.1, switching the roles of Y and W,

$$\sqrt{n}(\tilde{r}_C - \rho_C) \rightsquigarrow 12\,[Z_{1,\rho} - \rho_C Z_{3,\rho}] \text{ in } \mathbb{R},$$

where $Z_{3,\rho}$ is a zero-mean Gaussian random variable defined in Theorem 6.1.

Comment A.2 (Correlation-based vs. Regression-based Estimators)**.** The correlation-based estimators are asymptotically equivalent to the average of the regression-based and reversed regression-based restricted estimators. To see this equivalence, we combine $\sqrt{n}(\tilde{\varrho}_C - \rho_C) \rightsquigarrow 12\,[Z_{1,\rho} - \rho_C Z_{2,\rho}]$ with $\sqrt{n}(\tilde{r}_C - \rho_C) \rightsquigarrow 12\,[Z_{1,\rho} - \rho_C Z_{3,\rho}]$ to get

$$\sqrt{n}\,((\tilde{\varrho}_C + \tilde{r}_C)/2 - \rho_C) \rightsquigarrow 12\,[Z_{1,\rho} - \rho_C(Z_{2,\rho} + Z_{3,\rho})/2] \text{ in } \mathbb{R}.$$

The same result applies for the average of the regression-based and reversed regression-based unrestricted estimators.

Comment A.3 (Relative Efficiency)**.** The relative asymptotic efficiency of the different estimators depends on the variances of the components of the limit processes and the correlations between them. For example, the fully-restricted estimator $\breve{\rho}_C$ is relatively more efficient than the regression-based restricted estimator $\tilde{\varrho}_C$ if

$$\mathrm{Cor}(Z_{1,\rho}, \rho Z_{2,\rho}) \leqslant \frac{1}{2}\sqrt{\frac{\mathrm{Var}(\rho Z_{2,\rho})}{\mathrm{Var}(Z_{1,\rho})}},$$

and relative to the correlation-based estimator $\tilde{\rho}_C$ if

$$\mathrm{Cor}(Z_{1,\rho}, \rho(Z_{2,\rho} + Z_{3,\rho})/2) \leqslant \frac{1}{2}\sqrt{\frac{\mathrm{Var}(\rho(Z_{2,\rho} + Z_{3,\rho})/2)}{\mathrm{Var}(Z_{1,\rho})}},$$

The correlation-based estimator is relatively more efficient than the regression-based restricted estimator if $\mathrm{Var}(Z_{2,\rho}) = \mathrm{Var}(Z_{3,\rho})$ and $\mathrm{Cov}(Z_{1,\rho}, Z_{2,\rho}) = \mathrm{Cov}(Z_{1,\rho}, Z_{3,\rho})$.

Appendix B. Proofs of Section 6

B.1. **Hadamard Differentiability of CRRR Functionals.** We start by establishing the Hadamard differentiability of the fully-restricted functional ϕ_1 defined in (6.2) and characterizing the expression of the corresponding derivative. Next, we establish the Hadamard differentiability of the correlation-based and regression-based functionals ϕ and φ defined in (6.1) and (6.3), respectively, and characterize the corresponding derivatives. We provide a brief proof for the results for ϕ and φ because they follow by similar arguments as the proof for ϕ_1.

We need some setup and preliminary observations. For $\mathcal{R} \in \{\mathcal{Y}, \mathcal{W}\}$, let $\ell^\infty_m(\mathcal{RX})$ denote the set of all bounded and measurable mappings $\mathcal{RX} \mapsto \mathbb{R}$. Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ be the extended real line. We consider $\mathcal{RX}$ as a subset of $\overline{\mathbb{R}}^{d_x+1}$, with relative topology. Let ρ denote a standard metric on $\overline{\mathbb{R}}^{d_x+1}$. The closure of $\mathcal{RX}$ under ρ, denoted $\overline{\mathcal{RX}}$, is compact in $\overline{\mathbb{R}}^{d_x+1}$. Let $UC(\mathcal{RX}, \rho)$ be the set of functions mapping $\mathcal{RX}$ to the real line that are uniformly continuous with respect to the metric ρ , and can be continuously extended to $\overline{\mathcal{RX}}$, so that $UC(\mathcal{RX}, \rho) \subset \ell^\infty_m(\mathcal{RX})$. For a class of functions $\mathcal{F}$, let $UC(\mathcal{F},\lambda)$ be the set of functionals mapping $\mathcal{F}$ to the real line that are uniformly continuous with respect to the (semi) metric $\lambda(f,\tilde{f}) = [\mathrm{P}(f-\tilde{f})^2]^{1/2}$.

Lemma B.1 (Hadamard differentiability of ϕ_1). *Let $\mathcal{RX} \subseteq \mathbb{R}^{d_x+1}$, $\mathcal{R} \in \{\mathcal{Y},\mathcal{W}\}$, and $\mathcal{F}$ be the class of bounded functions, mapping $\overline{\mathbb{R}}^{d_x+2}$ to $\mathbb{R}$, that contains $F_{Y|X}$, $F_{W|X}$, $F_{Y|X}F_{W|X}$ and the indicators of all the rectangles in $\overline{\mathbb{R}}^{d_x+2}$, such that $\mathcal{F}$ is totally bounded under λ. Let $\mathbb{D}_\phi$ be the product of the spaces of measurable functions $\Gamma_Y : \mathcal{YX} \mapsto [-.5,.5]$ defined by $(y,x) \mapsto \Gamma_Y(y,x)$ and $\Gamma_W : \mathcal{WX} \mapsto [-.5,.5]$ defined by $(w,x) \mapsto \Gamma_W(w,x)$, and the bounded maps $\Pi : \mathcal{F} \mapsto \mathbb{R}$ defined by $f \mapsto \int f \mathrm{d}\Pi$, where Π is restricted to be a probability measure on $\mathcal{Z} := \mathcal{YWX}$. Consider the map $\phi_1 : \mathbb{D}_{\phi_1} \subset \mathbb{D} = \ell^\infty_m(\mathcal{Y}) \times \ell^\infty_m(\mathcal{WX}) \times \ell^\infty(\mathcal{F}) \to \mathbb{E} \subset \mathbb{R}$, defined by*

$$(\Gamma_Y,\Gamma_W,\Pi) \mapsto \phi_1(\Gamma_Y,\Gamma_W,\Pi) := \int \Gamma_Y(y,x)\Gamma_W(w,x)\mathrm{d}\Pi(z).$$

Then the map ϕ_1 is well defined. Moreover, the map ϕ_1 is Hadamard-differentiable at $(\Gamma_Y,\Gamma_W,\Pi) = (F_{Y|X}-.5, F_{W|X}-.5, F_Z)$, tangentially to the subset $\mathbb{D}_0 = UC(\mathcal{YX},\rho) \times UC(\mathcal{WX},\rho) \times UC(\mathcal{F},\lambda)$, with the derivative map $(\gamma_Y,\gamma_W,\pi) \mapsto \phi'_{1,F_{Y|X},F_{Y|X},F_Z}(\gamma_Y,\gamma_W,\pi)$ mapping $\mathbb{D}$ to $\mathbb{E}$ defined by

$$\begin{aligned}\phi'_{1,F_{Y|X},F_{W|X},F_Z}(\gamma_Y,\gamma_W,\pi) := &\int \gamma_Y(y,x)[F_{W|X}(w\mid x)-1/2]\mathrm{d}F_Z(y,w,x)\\ &+\int \gamma_W(w,x)[F_{Y|X}(y\mid x)-1/2]\mathrm{d}F_Z(y,w,x)\\ &+\int [F_{Y|X}(y\mid x)-1/2][F_{W|X}(w\mid x)-1/2]\mathrm{d}\pi(y,w,x),\end{aligned}$$

and the derivative is defined and is continuous on $\mathbb{D}$.

Proof of Lemma B.1. First we show that the map ϕ_1 is well defined. Any probability measure Π on $\mathcal{Z}$ is determined by the values $\int f\mathrm{d}\Pi$ for $f \in \mathcal{F}$, since $\mathcal{F}$ contains all the indicators of the rectangles in $\overline{\mathbb{R}}^{d_x+2}$. By Caratheodory's extension theorem $\Pi(A) = \Pi 1_A$ is well defined on all Borel subsets A of $\mathbb{R}^{d_x+2}$. Since $z \mapsto \Gamma_Y(y,x)\Gamma_W(y,x)$ is Borel measurable and takes values in $[-.5^2,.5^2]$, it follows that $\int \Gamma_Y(y,x)\Gamma_W(w,x)\mathrm{d}\Pi(z)$ is well defined as a Lebesgue integral, and $\int \Gamma_Y(y,x)\Gamma_W(w,x)\mathrm{d}\Pi(z) \in \mathbb{R}$.

Next we show the main claim. We establish the Hadamard differentiability of ϕ_1. Consider any sequence $(\Gamma^t_Y,\Gamma^t_W,\Pi^t) \in \mathbb{D}_\phi$ such that for $\gamma^t_Y := (\Gamma^t_Y - F_{Y|X}+.5)/t$, $\gamma^t_W := (\Gamma^t_W - F_{W|X}+.5)/t$, and $\pi^t(f) := \int f\mathrm{d}(\Pi^t - F_Z)/t$,

$$(\gamma^t_Y,\gamma^t_W,\pi^t) \to (\gamma_Y,\gamma_W,\pi), \text{ in } \ \ell^\infty_m(\mathcal{YX}) \times \ell^\infty_m(\mathcal{WX}) \times \ell^\infty(\mathcal{F}), \text{ where } (\gamma_Y,\gamma_W,\pi) \in \mathbb{D}_0.$$

We want to show that as $t \searrow 0$

$$\frac{\phi_1(\Gamma^t_Y,\Gamma^t_W,\Pi^t) - \phi_1(F_{Y|X}-.5,F_{W|X}-.5,F_Z)}{t} - \phi'_{1,F_{Y|X},F_{W|X},F_Z}(\gamma_Y,\gamma_W,\pi) \to 0 \text{ in } \mathbb{R}.$$

Write the difference above as

$$\int(\gamma_Y^t-\gamma_Y)[F_{W|X}-.5]\mathrm{d}F_Z+\int(\gamma_W^t-\gamma_W)[F_{Y|X}-.5]\mathrm{d}F_Z+\int[F_{Y|X}-.5][F_{W|X}-.5](\mathrm{d}\pi^t-\mathrm{d}\pi)$$
$$+\int\gamma_Y[F_{W|X}-.5]t\mathrm{d}\pi^t+\int\gamma_W[F_{Y|X}-.5]t\mathrm{d}\pi^t+\int(\gamma_Y^t-\gamma_Y)[F_{W|X}-.5]t\mathrm{d}\pi^t$$
$$+\int(\gamma_W^t-\gamma_W)[F_{Y|X}-.5]t\mathrm{d}\pi^t+\int\gamma_Y^t\gamma_W^t t\mathrm{d}F_Z+\int\gamma_Y^t\gamma_W^t t^2\mathrm{d}\pi^t \quad \text{(B.1)}$$

The first two terms of (B.1) are bounded by $\|\gamma_R^t-\gamma_R\|_{\mathcal{RX}}\int\mathrm{d}F_Z\to 0$, $R\in\{Y,W\}$. The third term vanishes, since for any $f\in\mathcal{F}$, $\int f\mathrm{d}\pi^t\to\int f\mathrm{d}\pi$ in $\ell^\infty(\mathcal{F})$, and $[F_{Y|X}-.5][F_{W|X}-.5]\in\mathcal{F}$ by assumption. The fourth and fifth terms vanish by the argument provided below. The sixth and seventh term vanish, since $|\int(\gamma_R^t-\gamma_R)t\mathrm{d}\pi^t|\leqslant\|\gamma_R^t-\gamma_R\|_{\mathcal{RX}}\int|t\mathrm{d}\pi^t|\leqslant 2\|\gamma_R^t-\gamma_R\|_{\mathcal{RX}}\to 0$ for $R\in\{Y,W\}$, where $\int|\mathrm{d}\mu|$ is the total variation of the signed measure μ. The eighth term is bounded by $Ct\int\mathrm{d}F_Z\to 0$, for some $C>0$. The last term can be bounded as:

$$2t\|\gamma_Y^t\|_{\mathcal{YX}}\|\gamma_W^t\|_{\mathcal{WX}}=2t\{\|\gamma_Y\|_{\mathcal{YX}}+o(1)\}\{(\gamma_W\|_{\mathcal{WX}}+o(1)\}\to 0.$$

Here we consider the fourth term $\int\gamma_Y[F_{W|X}-.5]t\mathrm{d}\pi^t$ and show that it vanishes. The argument for the fifth term is analogous. Since γ_Y is continuous on the compact semi-metric space $(\overline{\mathcal{YX}},\rho)$, there exists a finite partition of $\overline{\mathbb{R}}^{d_x+1}$ into non-overlapping rectangular regions $(R_{im}:1\leqslant i\leqslant m)$ (rectangles are allowed not to include their sides to make them non-overlapping) such that γ_Y varies at most ϵ on $\mathcal{YX}\cap R_{im}$. Let $p_m(y,x):=(y_{im},x_{im})$ if $(y,x)\in\mathcal{YX}\cap R_{im}$, where (y_{im},x_{im}) is an arbitrarily chosen point within $\mathcal{YX}\cap R_{im}$ for each i; also let $\chi_{im}(z):=1\{(y,x)\in R_{im}\}$. Then, as $t\to 0$,

$$\begin{aligned}\left|\int\gamma_Y[F_{W|X}-.5]td\pi^t\right|&\leqslant\left|\int(\gamma_Y-\gamma_Y\circ p_m)td\pi^t\right|+\left|\int(\gamma_Y\circ p_m)td\pi^t\right|\\&\leqslant\|\gamma_Y-\gamma_Y\circ p_m\|_{\mathcal{YX}}\int|td\pi^t|+\sum_{i=1}^m|\gamma_Y(y_{im},x_{im})|t\left|\pi^t(\chi_{im})\right|\\&\leqslant 2\|\gamma_Y-\gamma_Y\circ p_m\|_{\mathcal{YX}}+tm\|\gamma_Y\|_{\mathcal{YX}}\max_{1\leqslant i\leqslant m}\left|\pi^t(\chi_{im})\right|\\&\leqslant 2\epsilon+tm\|\gamma_Y\|_{\mathcal{YX}}\left\|\pi^t\right\|_{\mathcal{F}}\leqslant 2\epsilon+tm\left[\|\gamma_Y\|_{\mathcal{YX}}\|\pi\|_{\mathcal{F}}+o(1)\right]\leqslant 2\epsilon+O(t)\to 2\epsilon,\end{aligned}$$

since $\|F_{W|X}-.5\|_{\mathcal{WX}}<1$ and $\{\chi_{im}:1\leqslant i\leqslant m\}\subset\mathcal{F}$, so that $\max_i|\pi^t(\chi_{im})|\leqslant\|\pi^t\|_{\mathcal{F}}\to\|\pi\|_{\mathcal{F}}<\infty$.[17] The constant ϵ is arbitrary, so that the right hand side vanishes as $t\to 0$.

The derivative is well-defined over the entire $\mathbb{D}$ and is in fact continuous with respect to the norm on $\mathbb{D}$ given by $\|\cdot\|_{\mathcal{YX}}\vee\|\cdot\|_{\mathcal{WX}}\vee\|\cdot\|_{\mathcal{F}}$. The third component of the derivative map is trivially continuous with respect to $\|\cdot\|_{\mathcal{F}}$. The first component is continuous with respect to $\|\cdot\|_{\mathcal{YX}}$ since

$$\left|\int(\gamma_Y-\tilde{\gamma}_Y)[F_{W|X}-.5]\mathrm{d}F_Z(z)\right|\leqslant\|\gamma_Y-\tilde{\gamma}_Y\|_{\mathcal{YX}}\int\mathrm{d}F_Z(z).$$

[17] The set $\mathcal{F}$ is allowed to include zero, the indicator of an empty rectangle.

The second component is continuous with respect to $\|\cdot\|_{\mathcal{WX}}$ by an analogous argument. Hence the derivative map is continuous. ∎

Lemma B.2 (Hadamard differentiability of ϕ). *Let $\mathcal{RX} \subseteq \mathbb{R}^{d_x+1}$, $\mathcal{R} \in \{\mathcal{Y}, \mathcal{W}\}$, and $\mathcal{F}$ be the class of bounded functions, mapping $\overline{\mathbb{R}}^{d_x+2}$ to $\mathbb{R}$, that contains $F_{Y|X}$, $F_{W|X}$, $F^2_{W|X}$, $F^2_{Y|X}$, $F_{Y|X}F_{W|X}$ and the indicators of all the rectangles in $\overline{\mathbb{R}}^{d_x+2}$, such that $\mathcal{F}$ is totally bounded under λ. Let $\mathbb{D}_\phi$ be the product of the spaces of measurable functions $\Gamma_Y : \mathcal{YX} \mapsto [-.5, .5]$ defined by $(y, x) \mapsto \Gamma_Y(y, x)$ and $\Gamma_W : \mathcal{WX} \mapsto [-.5, .5]$ defined by $(w, x) \mapsto \Gamma_W(w, x)$, and the bounded maps $\Pi : \mathcal{F} \mapsto \mathbb{R}$ defined by $f \mapsto \int f \mathrm{d}\Pi$, where Π is restricted to be a probability measure on $\mathcal{YWX}$, $\int \Gamma_Y(y, x)^2 \mathrm{d}\Pi > 0$ and $\int \Gamma_W(w, x)^2 \mathrm{d}\Pi > 0$. Consider the map $\phi : \mathbb{D}_\phi \subset \mathbb{D} = \ell^\infty_m(\mathcal{Y}) \times \ell^\infty_m(\mathcal{WX}) \times \ell^\infty(\mathcal{F}) \to \mathbb{E} \subset \mathbb{R}$, defined by*

$$(\Gamma_Y, \Gamma_W, \Pi) \mapsto \phi(\Gamma_Y, \Gamma_W, \Pi) := \frac{\int \Gamma_Y(y, x)\Gamma_W(w, x)\mathrm{d}\Pi(z)}{\sqrt{\int \Gamma_W(w, x)^2 \mathrm{d}\Pi(z) \int \Gamma_Y(y, x)^2 \mathrm{d}\Pi(z)}}.$$

Then the map ϕ is well defined. Moreover, the map ϕ is Hadamard-differentiable at $(\Gamma_Y, \Gamma_W, \Pi) = (F_{Y|X} - .5, F_{W|X} - .5, F_Z)$, tangentially to the subset $\mathbb{D}_0 = UC(\mathcal{YX}, \rho) \times UC(\mathcal{WX}, \rho) \times UC(\mathcal{F}, \lambda)$, with the derivative map $(\gamma_Y, \gamma_W, \pi) \mapsto \phi'_{F_{Y|X}, F_{Y|X}, F_Z}(\gamma_Y, \gamma_W, \pi)$ mapping $\mathbb{D}$ to $\mathbb{E}$ defined by

$$\begin{aligned}&\phi'_{F_{Y|X}, F_{W|X}, F_Z}(\gamma_Y, \gamma_W, \pi) \\ &\qquad := 12[\phi'_{1, F_{Y|X}, F_{W|X}, F_Z}(\gamma_Y, \gamma_W, \pi) - \rho_C(\phi'_{2, F_{W|X}, F_Z}(\gamma_W, \pi) + \phi'_{3, F_{W|X}, F_Z}(\gamma_W, \pi))/2],\end{aligned}$$

with $\phi'_{1, F_{Y|X}, F_{W|X}, F_Z}(\gamma_Y, \gamma_W, \pi)$ defined as in Lemma B.1,

$$\begin{aligned}\phi'_{2, F_{W|X}, F_Z}(\gamma_W, \pi) := 2 \int \gamma_W(w, x)[F_{W|X}(w \mid x) - .5]\mathrm{d}F_Z(y, w, x) \\ + \int [F_{W|X}(w \mid x) - .5]^2 \mathrm{d}\pi(y, w, x),\end{aligned}$$

and

$$\begin{aligned}\phi'_{3, F_{Y|X}, F_Z}(\gamma_Y, \pi) := 2 \int \gamma_Y(y, x)[F_{Y|X}(y \mid x) - .5]\mathrm{d}F_Z(y, w, x) \\ + \int [F_{Y|X}(y \mid x) - .5]^2 \mathrm{d}\pi(y, w, x);\end{aligned}$$

where the derivative is defined and is continuous on $\mathbb{D}$.

Proof of Lemma B.2. It is convenient to express

$$\phi(\Gamma_Y, \Gamma_W, \Pi) = \frac{\phi_1(\Gamma_Y, \Gamma_W, \Pi)}{\sqrt{\phi_2(\Gamma_W, \Pi)\phi_3(\Gamma_Y, \Pi)}},$$

where ϕ_1 defined as in Lemma B.1,

$$\phi_2(\Gamma_W, \Pi) := \int \Gamma_W(w, x)^2 \mathrm{d}\Pi(z) \quad \text{and} \quad \phi_3(\Gamma_Y, \Pi) := \int \Gamma_Y(y, x)^2 \mathrm{d}\Pi(z).$$

First note that the maps ϕ_2 and ϕ_3 are well defined by a similar argument to the proof of Lemma B.1 that shows that ϕ_1 is well-defined. The map ϕ is also well-defined because $\phi_2(\Gamma_W, \Pi) > 0$ and $\phi_3(\Gamma_Y, \Pi) > 0$ by assumption.

Next we show the main claim. The Hadamard differentiability of of ϕ_1 is establish in Lemma B.1. The Hadamard differentiability of ϕ_2 and ϕ_3 can be established by analogous arguments. In particular, the maps ϕ_2 and ϕ_3 in the denominator are Hadamard differentiable at $(F_{W|X} - .5, F_Z)$ and $(F_{Y|X} - .5, F_Z)$, respectively, with derivatives $\phi'_{2,F_{W|X},F_Z}$ and $\phi'_{3,F_{Y|X},F_Z}$. Indeed, we can show that as $t \searrow 0$

$$\frac{\phi_2(\Gamma_W^t, \Pi^t) - \phi_2(F_{W|X} - .5, F_Z)}{t} - \phi'_{2,F_{W|X},F_Z}(\gamma_W, \pi) \to 0 \text{ in } \mathbb{R}$$

and

$$\frac{\phi_3(\Gamma_Y^t, \Pi^t) - \phi_3(F_{Y|X} - .5, F_Z)}{t} - \phi'_{3,F_{Y|X},F_Z}(\gamma_Y, \pi) \to 0 \text{ in } \mathbb{R}$$

following an analogous argument as for ϕ_1 in the proof of Lemma B.1. It can also be showed that $\phi'_{2,F_{W|X},F_Z}$ and $\phi'_{3,F_{Y|X},F_Z}$ are well-defined over the entire $\mathbb{D}$ and are continuous. We omit the proof for the sake of brevity.

The final result then follows by the chain-rule for Hadamard differentiable maps using that

$$\mathrm{Var}(F_{W|X}(W \mid X)) = \mathrm{Var}(F_{Y|X}(Y \mid X)) = 1/12.$$

Continuity of the derivative with respect to the norm on $\mathbb{D}$ given by $\|\cdot\|_{\mathcal{YX}} \vee \|\cdot\|_{\mathcal{WX}} \vee \|\cdot\|_{\mathcal{F}}$ follows by continuity of $\phi'_{1,F_{Y|X},F_{W|X},F_Z}$, $\phi'_{2,F_{W|X},F_Z}$ and $\phi'_{3,F_{Y|X},F_Z}$. ∎

Lemma B.3 (Hadamard differentiability of φ)**.** *Let $\mathcal{RX} \subseteq \mathbb{R}^{d_x+1}$, $\mathcal{R} \in \{\mathcal{Y}, \mathcal{W}\}$, and $\mathcal{F}$ be the class of bounded functions, mapping $\overline{\mathbb{R}}^{d_x+2}$ to $\mathbb{R}$, that contains $F_{Y|X}$, $F_{W|X}$, $F^2_{W|X}$, $F_{Y|X}F_{W|X}$ and the indicators of all the rectangles in $\overline{\mathbb{R}}^{d_x+2}$, such that $\mathcal{F}$ is totally bounded under λ. Let $\mathbb{D}_\phi$ be the product of the spaces of measurable functions $\Gamma_Y : \mathcal{YX} \mapsto [-.5, .5]$ defined by $(y, x) \mapsto \Gamma_Y(y, x)$ and $\Gamma_W : \mathcal{WX} \mapsto [-.5, .5]$ defined by $(w, x) \mapsto \Gamma_W(w, x)$, and the bounded maps $\Pi : \mathcal{F} \mapsto \mathbb{R}$ defined by $f \mapsto \int f \mathrm{d}\Pi$, where Π is restricted to be a probability measure on $\mathcal{YWX}$ and $\int \Gamma_W(w, x)^2 \mathrm{d}\Pi > 0$. Consider the map $\varphi : \mathbb{D}_\phi \subset \mathbb{D} = \ell^\infty_m(\mathcal{Y}) \times \ell^\infty_m(\mathcal{WX}) \times \ell^\infty(\mathcal{F}) \to \mathbb{E} \subset \mathbb{R}$, defined by*

$$(\Gamma_Y, \Gamma_W, \Pi) \mapsto \varphi(\Gamma_Y, \Gamma_W, \Pi) := \frac{\int \Gamma_Y(y, x)\Gamma_W(w, x)\mathrm{d}\Pi(z)}{\int \Gamma_W(w, x)^2 \mathrm{d}\Pi(z)}.$$

Then the map φ is well defined. Moreover, the map φ is Hadamard-differentiable at $(\Gamma_Y, \Gamma_W, \Pi) = (F_{Y|X} - .5, F_{W|X} - .5, F_Z)$, tangentially to the subset $\mathbb{D}_0 = UC(\mathcal{YX}, \rho) \times UC(\mathcal{WX}, \rho) \times UC(\mathcal{F}, \lambda)$, with the derivative map $(\gamma_Y, \gamma_W, \pi) \mapsto \varphi'_{F_{Y|X},F_{Y|X},F_Z}(\gamma_Y, \gamma_W, \pi)$ mapping $\mathbb{D}$ to $\mathbb{E}$ defined by

$$\varphi'_{F_{Y|X},F_{W|X},F_Z}(\gamma_Y, \gamma_W, \pi) := 12[\phi'_{1,F_{Y|X},F_{W|X},F_Z}(\gamma_Y, \gamma_W, \pi) - \rho_C \phi'_{2,F_{W|X},F_Z}(\gamma_W, \pi)],$$

with $\phi'_{1,F_{Y|X},F_{W|X},F_Z}(\gamma_Y, \gamma_W, \pi)$ defined as in Lemma B.1 and $\phi'_{2,F_{W|X},F_Z}(\gamma_W, \pi)$ defined as in Lemma B.2, where the derivative is defined and is continuous on $\mathbb{D}$.

Proof of Lemma B.3. The result follows by an analogous argument to the proof of Lemma B.2. We omit the proof for the sake of brevity. ∎

B.2. **Proof of Lemma 6.1.** We start by stating a Lemma with a bootstrap functional central limit theorem for the bootstrap draws of the inputs needed to establish Theorem 6.2. We shall prove this lemma together with Lemma 6.1.

For $R \in \{Y, W\}$, let $(r, x) \mapsto \widehat{Z}^*_R(r, x) := \sqrt{n}\left(\widehat{F}^*_{R|X}(r \mid x) - \widehat{F}_{R|X}(r \mid x)\right)$ and $f \mapsto \widehat{G}^*_Z(f) := \sqrt{n}\int f \mathrm{d}(\widehat{F}^*_Z - \widehat{F}_Z)$, where $\widehat{F}^*_{R|X}(r \mid x) := \Lambda(x'\widehat{\beta}^*_R(r))$, $\widehat{\beta}^*_R(r)$ is the bootstrap draw of $\widehat{\beta}_R(r)$ defined in Algorithm 2 and $\widehat{F}^*_Z$ is the bootstrap draw of the empirical distribution function of Z, be exchangeable bootstrap draws of the empirical processes $(r, x) \mapsto \widehat{Z}_R(r, x)$ and $f \mapsto \widehat{G}_Z(f)$.

Lemma B.4 (Bootstrap Limit Processses for Inputs)**.** *Under the conditions of Lemma 6.1 and Assumption 6.2, in the metric space $\ell^\infty(\mathcal{YWXF})$,*

$$(\widehat{Z}^*_Y(y, x), \widehat{Z}^*_W(w, x), \widehat{G}^*_Z(f)) \rightsquigarrow_{\mathrm{P}} (Z_Y(y, x), Z_W(w, x), G_Z(f)),$$

as stochastic processes indexed by (y, w, x, f), where $(Z_Y(y, x), Z_W(w, x), G_Z(f))$ has the same distribution as the limit process in Lemma 6.1.

The proof of Lemmas 6.1 and B.4 follows similar steps to the proof of Theorem 5.2 in Chernozhukov et al. (2013), suitably modified to extend the process to the tails. The main differences are highlighted in Steps 1, 2, and 3 below.

Step 1.(Results for coefficients and empirical measures). Application of the Hadamard differentiability results for Z-processes in Chernozhukov et al. (2013) gives that, in $\ell^\infty(\bar{\mathcal{Y}})^{d_x} \times \ell^\infty(\bar{\mathcal{W}})^{d_x} \times \ell^\infty(\mathcal{F})$,

$$(\sqrt{n}(\widehat{\beta}_Y(\cdot) - \beta_Y(\cdot)), \sqrt{n}(\widehat{\beta}_W(\cdot) - \beta_W(\cdot)), \widehat{G}_Z) \rightsquigarrow (H_Y(\cdot), H_W(\cdot), G_Z), \tag{B.2}$$

where $\bar{\mathcal{Y}}$ and $\bar{\mathcal{W}}$ are any compact strict subsets of $\mathcal{Y}$ and $\mathcal{W}$, respectively, and

$$r \mapsto H_R(r) := -J_R(r)^{-1}\mathbb{G}(\varphi_{r,\beta}), \quad \varphi_{r,\beta}(R, X) := \left[\Lambda(X'\beta_R(r)) - 1\{R \leqslant r\}\right] X,$$

has continuous paths a.s., for $R \in \{Y, W\}$.[18]

We extend the process $\widehat{\beta}(r)$ to the tails as

$$\widehat{\beta}_R(r) = \widehat{\beta}_R(\bar{r}) + (r - \bar{r})\widehat{\alpha}_R(\bar{r})e_1, \quad r \in \mathcal{R} \setminus \bar{\mathcal{R}},$$

where e_1 is a unitary d_x-vector with a one in the first component. Likewise, the estimands are given by

$$\beta_R(r) = \beta_R(\bar{r}) + (r - \bar{r})\alpha_R(\bar{r})e_1 \quad r \in \mathcal{R} \setminus \bar{\mathcal{R}},$$

by assumption.

In what follows it is convenient to analyze the estimator for the lower tail, the analysis for estimators for upper tails follows exactly the same steps, switching the signs on the dependent variables,

[18]Chernozhukov et al. (2013) gives detailed arguments on how H-differentiability of Z-processes implies that $(\sqrt{n}(\widehat{\beta}_Y(y) - \beta_Y(y)), \widehat{G}_X(f)) \rightsquigarrow (H_Y(y), G_X(f))$ in $\ell^\infty(\bar{\mathcal{Y}})^{d_x} \times \ell^\infty(\mathcal{F})$, where $\widehat{G}_X(f)$ is the empirical process induced by the marginal distribution of X. The extension to stacking another Z-process is straightforward, implying the result (B.2).

$R \in \{Y, W, -Y, -W\}$. The estimators $(\widehat{\beta}_R(\bar{r}), \widehat{\alpha}_R(\bar{r}))$ can be seen as Z-estimators with moment function

$$\bar{\varphi}_{\beta,\alpha}(R, X) = \left(\varphi_{\bar{r},\beta}(R, X)', \varphi_\alpha(R, X)(r_0 - \bar{r})\right)', \ \varphi_\alpha(R, X) := \Lambda(X'\beta_R(\bar{r}) + (r_0 - \bar{r})\alpha_R(\bar{r})) - 1\{R \leqslant r_0\}.$$

Invoking Z-process theory again but this time for the simple case of finite-dimensional space $\mathbb{R}^{d_x+1}$, we have that jointly in $R \in \{Y, -Y, W, -W\}$,

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\beta}_R(\bar{r}) - \beta_R(\bar{r}), \widehat{\alpha}_R(\bar{r}) - \alpha_R(\bar{r})\right)' \\
&\quad \rightsquigarrow \begin{bmatrix} J_R(\bar{r}) & 0 \\ (r_0 - \bar{r})\mathrm{E}[\lambda(X'\beta_R(r_0))X'] & (r_0 - \bar{r})^2\mathrm{E}[\lambda(X'\beta_R(r_0))] \end{bmatrix}^{-1} \mathbb{G}(\bar{\varphi}_{\beta,\alpha}) \\
&\quad = \begin{bmatrix} J_R^{-1}(\bar{r})\mathbb{G}(\varphi_{\bar{r},\beta}) \\ \frac{\mathbb{G}(\varphi_\alpha) - \mathrm{E}[\lambda(X'\beta_R(r_0))X]' J_R^{-1}(\bar{r})\mathbb{G}(\varphi_{\bar{r},\beta})}{(r_0 - \bar{r})\mathrm{E}[\lambda(X'\beta_R(r_0))]} \end{bmatrix}. \qquad \text{(B.3)}
\end{aligned}$$

In fact using Hadamard differentiability results for Z-processes given in Chernozhukov et al. (2013), we conclude that convergence results (B.2) and (B.3) for all $R \in \{Y, W, -Y, -W\}$ hold jointly.[19]

The Hadamard differentiability results for Z-processes also imply that the bootstrap analogs of the results (B.2) and (B.3) are also valid and hold jointly as well. We omit writing the formulas for these convergence results, since they are analogous to (B.2) and (B.3).

Step 2.(Main: Results for conditional cdfs). Here we show that,

$$\begin{aligned}
(\widehat{Z}_Y, \widehat{Z}_W, \widehat{G}_Z) &\rightsquigarrow (Z_Y, Z_W, G_Z) \text{ in } \ell^\infty(\mathcal{YWXF}), \\
(\widehat{Z}_Y^*, \widehat{Z}_W^*, \widehat{G}_Z^*) &\rightsquigarrow_{\mathrm{P}} (Z_Y, Z_W, G_Z) \text{ in } \ell^\infty(\mathcal{YWXF}).
\end{aligned}$$

For the body part, $r \in \overline{\mathcal{R}}$, consider the mapping $\nu : \mathbb{D}_\nu \subset \ell^\infty(\overline{\mathcal{R}}^{d_x}) \to \ell^\infty(\overline{\mathcal{R}}\mathcal{X})$, defined as

$$b \mapsto \nu(b), \quad \nu(b)(x, y) := \Lambda\left(x'b(y)\right).$$

It is straighforward to deduce that this map is Hadamard differentiable at $b(\cdot) = \beta_R(\cdot)$ tangentially to $UC(\mathcal{R}, \rho)^{d_x}$ with the derivative map given by:

$$v \mapsto \nu'(v), \quad \nu'(v)(r, x) = \lambda\left(x'\beta_R(r)\right) x'v(r).$$

For the tail part $r \in \mathcal{R} \setminus \overline{\mathcal{R}}$, we consider the mapping $\mu : \mathbb{R}^{d_x+1} \to \ell^\infty(\mathcal{R} \setminus \overline{\mathcal{R}})$ defined by:

$$(d, a) \mapsto \mu(d, a), \quad \mu(d, a)(r, x) := \Lambda(x'd + (r - \bar{r})ae_1)$$

This is also Hadamard differentiable at $(d, a) = (\beta_R(\bar{r}), \alpha_R(\bar{r}))$ tangentially to the entire domain with the derivative

$$(h, u) \mapsto \mu'(h, u), \quad \mu'(h, u)(r, x) = \lambda(x'\beta_R(r))(x'h + (r - \bar{r})ue_1).$$

[19] Of course, it is cumbersome to put this joint convergence statement into one display, so we state this verbally.

The derivative is a bounded (continuous) linear operator (note that as $r \to \pm\infty$, the derivative vanishes, with the linear growth factor $r - \bar{r}$ being dominated by the term $\lambda(x'\beta_R(r))$ with exponential tails).

We can now define the "extended map" that combines the body and tail pieces:

$$(b, d, a) \mapsto \bar{\nu}(b, d, a); \quad \bar{\nu}(b, d, a)(r, x) := \nu(b)(r, x)1(r \in \overline{\mathcal{R}}) + \mu(d, a)(r, x)1(r \in \mathcal{R} \setminus \overline{\mathcal{R}}).$$

By combining the two differentiability results above, we can deduce that this map is Hadamard differentiable with the derivative map

$$(v, h, u) \mapsto \bar{\nu}'(v, h, u), \quad \bar{\nu}'(v, h, u)(r, x) := \nu'(v)(r, x)1(r \in \overline{\mathcal{R}}) + \mu'(h, u)(r, x)1(r \in \mathcal{R} \setminus \overline{\mathcal{R}}).$$

Then, the claim follows by the functional delta method, and we find that the limit process is given by

$$Z_R(r, x) = \lambda(x'\beta_R(r))x'H_R(r), \quad R \in \{Y, W\}.$$

where we now define the extended version of H_R as: $H_R(r) = \mathbb{G}(\psi_{R,r,\beta_R})$ where

$$\psi_{R,r,\beta_R}(R, X) = 1\{r \in \bar{\mathcal{R}}\}J_R(r)^{-1}\varphi_{r,\beta}(R, X) + 1\{r \in \mathcal{R} \setminus \bar{\mathcal{R}}\}\Big[J_R(\bar{r})^{-1}\varphi_{\bar{r},\beta}(R, X) + \frac{r - \bar{r}}{r_0 - \bar{r}} \frac{\varphi_\alpha(R, X) - \mathrm{E}[\lambda(X'\beta_R(r_0))X]'J_R(\bar{r})^{-1}\varphi_{\bar{r},\beta}(R, X)}{\mathrm{E}[\lambda(X'\beta_R(r_0))]} e_1\Big],$$

and letting $\ell_{r,x}(R, X) := \lambda(x'\beta_R(r))x'\psi_{r,\beta_R}(R, X)$, after some algebra.

Step 3. (Auxiliary: Donskerness). One key ingredient for the result is to show that $\mathcal{F}$ is a Dudley-Koltchinskii-Pollard (DKP) class, namely it has bounded uniform covering entroy integral and obeys standard measurability condition (Dudley's image-admissible Suslin condition). We omit any discussion of measurability in this paper, but we note that it trivially holds. The proof in Chernozhukov et al. (2013) relies on compactness of the set $\bar{\mathcal{R}}\mathcal{X}$ and does not apply immediately. We extend the result to $\mathcal{R}\mathcal{X}$. For $R \in \{Y, W\}$, note that $\mathcal{F}_R = \{F_{R|X}(r \mid \cdot) : r \in \mathcal{R}\}$ is a uniformly bounded "parametric" family indexed by $r \in \mathcal{R}$ that obeys $|F_{R|X}(r \mid \cdot) - F_{R|X}(r' \mid \cdot)| \leqslant L|r - r'|$, given the assumption that the density function $f_{R|X}$ is uniformly bounded by some constant L. This was enough to bound the covering numbers for the index set $\bar{\mathcal{R}}$, but is not enough to bound the covering number over the unbounded set $\mathcal{R}$.

Under our modelling hypotheses, there exists a small enough constant $C > 0$ such that

$$F_{R|X}(r \mid \cdot) \leqslant \exp(rC); \;\; r < 0; \quad 1 - F_{R|X}(r \mid \cdot) \leqslant \exp(-rC); \;\; r > 0;$$

Let $R_j = -M(\epsilon) - \epsilon/(2L) + j(\epsilon/L)$, with $j = 0, ..., J$, where $J = \lceil 2M(\epsilon)L/\epsilon \rceil + 1$, $M(\epsilon) = \log(1/\epsilon)/C$ and $\lceil \cdot \rceil$ is the ceiling function. Let $R_{-1} = -\infty$ and $R_{J+1} = +\infty$. The sets $B_j = \{F_{R|X}(r \mid X) : R_j \leqslant r \leqslant R_{j+1}\}$ for $j \in \{-1, ..., J\}$ have the L^2 diameter of at most ϵ independently of the distribution of F_X:

- Indeed by the previous paragraph, if $j \in \{0, ..., J - 1\}$, then the diameter of the set B_j is at most $L(\epsilon/L) = \epsilon$.

- For $j = -1$ or J, then any pair of conditional cdfs in the same ball obey:

$$|F_{R|X}(r \mid \cdot) - F_{R|X}(r' \mid \cdot)| \leqslant \exp(-M(\epsilon)C) = \exp(-[\log(1/\epsilon)/C]C) \leqslant \epsilon.$$

The number of sets is at most $2\log(1/\epsilon)L/(C\epsilon) + 5$. It follows that the uniform covering entropy of the function set $\mathcal{F}_R = \{F_{R|X}(r \mid X) : r \in \mathcal{R}\}$ is bounded by $(1 + (1/\epsilon)^2)$ up to a constant that does not depend on the distribution of X.

Further if we take $\mathcal{F}$ as generated by union of products of $\mathcal{F}_R$ over different labels R, and the union of rectangles, the resulting set is still a DKP class, by standard uniform covering entropy calculus. ■

B.3. **Proof of Theorem 6.1.** Part (1) follows by the functional delta method (see, e.g., Lemma B.1 of Chernozhukov et al. (2013)). Indeed, in the notation of Lemma B.2,

$$\tilde{\rho}_C = \phi(\widehat{F}_{Y|X}, \widehat{F}_{W|X}, \widehat{F}_Z) = \frac{\int[\widehat{F}_{Y|X}(y \mid x) - .5][\widehat{F}_{W|X}(W \mid x) - .5]\mathrm{d}\widehat{F}_Z(y,w,x)}{\sqrt{\int[\widehat{F}_{W|X}(W \mid x) - .5]^2\mathrm{d}\widehat{F}_Z(y,w,x)\int[\widehat{F}_{Y|X}(Y \mid x) - .5]^2\mathrm{d}\widehat{F}_Z(y,w,x)}},$$

and $\rho_C = \phi(F_{Y|X}, F_{W|X}, F_Z)$. By Lemmas B.2 and 6.1, together with the functional delta method,

$$\sqrt{n}(\tilde{\rho}_C - \rho_C) \rightsquigarrow \phi'_{F_{Y|X},F_{W|X},F_Z}(Z_Y, Z_W, G_Z) \text{ in } \mathbb{R}.$$

Similarly, in the notation of Lemma B.1,

$$\breve{\rho}_C = 12\phi_1(\widehat{F}_{Y|X}, \widehat{F}_{W|X}, \widehat{F}_Z) = 12\int[\widehat{F}_{Y|X}(y \mid x) - .5][\widehat{F}_{W|X}(W \mid x) - .5]\mathrm{d}\widehat{F}_Z(y,w,x),$$

and $\rho_C = 12\phi_1(F_{Y|X}, F_{W|X}, F_Z)$. By Lemmas B.1 and 6.1, together with the functional delta method,

$$\sqrt{n}(\breve{\rho}_C - \rho_C) \rightsquigarrow 12\,\phi'_{1,F_{Y|X},F_{W|X},F_Z}(Z_Y, Z_W, G_Z) \text{ in } \mathbb{R}.$$

To show part (2), write

$$\frac{1}{n}\sum_{i=1}^n(\widehat{V}_i - \overline{\widehat{V}})^2 = \frac{1}{n}\sum_{i=1}^n(\widehat{V}_i - .5)^2 - (\overline{\widehat{V}} - .5)^2, \quad \overline{\widehat{V}} = \frac{1}{n}\sum_{i=1}^n\widehat{V}_i,$$

$$\frac{1}{n}\sum_{i=1}^n(\widehat{U}_i - \overline{\widehat{U}})^2 = \frac{1}{n}\sum_{i=1}^n(\widehat{U}_i - .5)^2 - (\overline{\widehat{U}} - .5)^2, \quad \overline{\widehat{U}} = \frac{1}{n}\sum_{i=1}^n\widehat{U}_i,$$

and

$$\frac{1}{n}\sum_{i=1}^n\widehat{U}_i(\widehat{V}_i - \overline{\widehat{V}}) = \frac{1}{n}\sum_{i=1}^n(\widehat{U}_i - .5)(\widehat{V}_i - .5) - (\overline{\widehat{U}} - .5)(\overline{\widehat{V}} - .5).$$

To analyze the second components of the previous expressions, note that

$$|\overline{\widehat{V}} - .5| \leqslant \frac{1}{n}\sum_{i=1}^n|\widehat{V}_i - V_i| + \left|\frac{1}{n}\sum_{i=1}^n(V_i - .5)\right| = O_{\mathrm{P}}(n^{-1/2}), \quad V_i = F_{W|X}(W_i \mid X_i),$$

by the central limit theorem applied to the second term and

$$\max_{i\in\{1,\ldots,n\}} |\widehat{V}_i - V_i| = \max_{i\in\{1,\ldots,n\}} |\widehat{F}_{W|X}(W_i \mid X_i) - F_{W|X}(W_i \mid X_i)| \leqslant \|\widehat{F}_{W|X} - F_{W|X}\|_{\mathcal{W}\mathcal{X}} = O_{\mathrm{P}}(n^{-1/2}), \tag{B.4}$$

where the last equality follows by Lemma 6.1. A similar argument gives $|\overline{\widehat{U}} - .5| = O_{\mathrm{P}}(n^{-1/2})$, and

$$\max_{i\in\{1,\ldots,n\}} |\widehat{U}_i - U_i| = O_{\mathrm{P}}(n^{-1/2}), \quad U_i = F_{Y|X}(Y_i \mid X_i). \tag{B.5}$$

Combining the previous results with the continuous mapping theorem, and using a mean value expansion, we conclude that

$$\widehat{\rho}_C = \frac{\frac{1}{n}\sum_{i=1}^n (\widehat{U}_i - .5)(\widehat{V}_i - .5) + O_{\mathrm{P}}(n^{-1})}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^n (\widehat{V}_i - .5)^2 + O_{\mathrm{P}}(n^{-1})\right]\left[\frac{1}{n}\sum_{i=1}^n (\widehat{U}_i - .5)^2 + O_{\mathrm{P}}(n^{-1})\right]}} = \tilde{\rho}_C + O_{\mathrm{P}}(n^{-1}).$$

■

B.4. **Proof of Theorem 6.2.** The results for $\tilde{\rho}_C^*$ and $\breve{\rho}_C^*$ follow by Lemmas B.2, B.1 and B.4, together with the functional delta method for bootstrap (see, e.g., Lemma B.3 of Chernozhukov et al. (2013)).

The result for $\widehat{\rho}_C^*$ follows from

$$\sqrt{n}(\widehat{\rho}_C^* - \widehat{\rho}_C) = \sqrt{n}(\tilde{\rho}_C^* - \tilde{\rho}_C) + \sqrt{n}(\tilde{\rho}_C - \widehat{\rho}_C) + \sqrt{n}(\widehat{\rho}_C^* - \tilde{\rho}_C^*) = \sqrt{n}(\tilde{\rho}_C^* - \tilde{\rho}_C) + o_{\mathrm{P}}(1),$$

because $\sqrt{n}(\tilde{\rho}_C - \widehat{\rho}_C) = o_{\mathrm{P}}(1)$ by part (2) of Theorem 6.1, and $\sqrt{n}(\widehat{\rho}_C^* - \tilde{\rho}_C^*) = o_{\mathrm{P}}(1)$ by the same argument as in the proof of part (2) of Theorem 6.1 replacing Lemma 6.1 by Lemma B.4.

Appendix C. Simulations

We show that the asymptotic theory provides a good approximation to the behavior of the CRRR estimator through a small sample Monte Carlo simulation. In particular, we document that the CRRR estimator converges at the expected rate and that the corresponding confidence interval has coverage close to its nominal level. We focus on the correlation-based estimator, but we find very similar performance for the regression-based and fully-restricted estimators in results not reported.[20]

We consider a bivariate normal design for analytical convenience. In particular, we draw data from the process

$$\begin{pmatrix} Y \\ W \end{pmatrix} \mid X = x \sim N_2\left(\begin{pmatrix} x \\ x \end{pmatrix}, \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}\right), \quad X \sim N(0,1). \tag{C.1}$$

We consider 3 different values for the correlation parameter $c \in \{0.25, 0.50, 0.75\}$ and 3 sample sizes $n \in \{625, 2500, 10000\}$. We choose these sample sizes because they correspond to $\sqrt{n} \in$

[20]The results for the regression-based and fully-restricted estimators are available from the authors upon request.

$\{25, 50, 100\}$, where $\sqrt{n}$ is the theoretical rate of convergence of the CRRR estimator. The true value of the CRRR slope is obtained from c using the expression of the rank correlation of the bivariate normal, $\rho_C = 6 \arcsin(c/2)/\pi$ (e.g., Cramér, 1999).

Table 8 shows the result from $2,000$ simulations using the correlation-based estimator of Algorithm 1 with Gaussian or probit link function, a mesh of 500 points located at sample quantiles in a sequence of orders from 0.005 to 0.995 with increments of 0.99/499, and $m = 30$ to estimate the tail parameters. We report root mean squared error (RMSE), bias, standard deviation (SD) and coverage of 95% confidence intervals (Cover.). The confidence intervals are obtained from Algorithm 3 by empirical bootstrap with 200 repetitions. The results indicate that (i) the estimator converges at the expected rate of $\sqrt{n}$ and (ii) the coverage of the confidence intervals floats around the nominal level of $.95$.[21]

Table 8. Properties of CRRR Estimator ($m = 30$)

c	n	RMSE	Bias	SD	Cover.
0.25	625	0.038	0.002	0.038	0.94
	2,500	0.020	0.001	0.020	0.94
	10,000	0.009	0.001	0.009	0.94
0.5	625	0.032	0.004	0.032	0.94
	2,500	0.016	0.002	0.016	0.94
	10,000	0.008	0.002	0.008	0.95
0.75	625	0.021	0.005	0.020	0.96
	2,500	0.010	0.002	0.010	0.95
	10,000	0.005	0.002	0.005	0.94

Notes: results based on 2,000 simulations of the DGP in (C.1) *for the correlation-based estimator with probit link function and a mesh of 500 points. The nominal level for coverage is* 0.95.

Table 9 report results using the same design as in table 8 but with $m = 0$, that is, without imposing any restriction at the tails. We find that the results are not sensitive to the treatment of the tails, even for the smallest sample size.

[21] Note that the simulation standard error for coverage is about 0.5%.

Table 9. Properties of CRRR Estimator ($m = 0$)

c	n	RMSE	Bias	SD	Cover.
0.25	625	0.038	0.003	0.038	0.95
	2,500	0.020	0.001	0.020	0.94
	10,000	0.009	0.001	0.009	0.94
0.5	625	0.032	0.005	0.032	0.95
	2,500	0.016	0.002	0.016	0.94
	10,000	0.008	0.001	0.008	0.95
0.75	625	0.022	0.007	0.020	0.96
	2,500	0.010	0.002	0.010	0.95
	10,000	0.005	0.001	0.005	0.94

Notes: results based on 1,500 simulations of the DGP in (C.1) *for the correlation-based estimator with probit link function and a mesh of 500 points. The nominal level for coverage is* 0.95.