UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 195

On predictions, measurement, and causal inference

Essays on statistical correlations and statistical applications in labor market economics and criminology

PETRA ORNSTEIN

ISSN 1652-9030
ISBN 978-91-513-1427-3
URN urn:nbn:se:uu:diva-467545

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Friday, 29 April 2022 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Kenneth Carling (Högskolan Dalarna).

Abstract

Ornstein, P. 2022. On predictions, measurement, and causal inference. Essays on statistical correlations and statistical applications in labor market economics and criminology. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 195. 26 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1427-3.

This thesis consists of five papers covering a broad range of statistical themes, from prediction via correlation measures to causal inference.

Paper I uses a nationally representative survey of separated Swedish women to examine whether men who strive to control their partners during their relationships are more likely to stalk or assault their ex-partners after separation. The empirical analysis shows that basic measures of control behaviors explain 18% of the variance in stalking victimization and 8% of the assault victimization. The predictive value of the constructed measures of control is shown to by far surpass those of other common risk markers included in the analysis.

Paper II presents asymptotic properties of Spearman's rank correlation for variables with finite support. For variables with finite support, the population version of Spearman's rank correlation has been derived. Using this result, convergence to a normal distribution irrespectively of dependence is shown, and the asymptotic variance is derived. A small simulation study indicates that the asymptotic properties are of practical importance.

Paper III proposes a robust version of the polychoric correlation, a parametric measure of the correlation between two unobserved continuous variables when the observed variables are discrete. Robust polychoric correlation is shown to be consistent and asymptotically normal. Results from a systematic Monte Carlo simulation suggest that the new estimator has better robustness properties than normality based maximum likelihood estimation of the polychoric correlation, to negligible efficiency costs.

Paper IV reports results from a large-scale randomized experiment evaluating whether a supported employment rehabilitation intervention strategy can improve labor market opportunities for young adults on disability pension better than regular vocational rehabilitation. In total, 1,062 individuals were randomly assigned between interventions. The main result shows that 18 months after the start of the project, participants offered supported employment have work rates that are approximately 10 percentage points higher than participants who received regular rehabilitation.

Paper V estimates consequences of interpersonal violence on victims. Using rich administrative population data for Sweden, individuals who visited a hospital in the years 1998 to 2002 due to assault are compared to individuals who did not have a hospital visit due to assault at that time, but who were statistically indistinguishable from the individuals of interest in the four years prior to the incident. The results suggest that violent crime has large and persistent effects on mortality, suicide, earnings, work status, disposable income, as well as on the number of days on sick leave. Specifically, an assault leading to a hospital visit is estimated to convey losses amounting to 1.4 million SEK per victimized woman and 1.5 million SEK per victimized man, whereof more than 80 percent result from excess mortality. Estimates on socio-economic outcomes show robustness against selection on unobserved characteristics. Estimates on mortality and suicides are very robust.

Keywords: risk marker studies, correlation measures, identification, field experiments, observational studies

Petra Ornstein, Department of Statistics, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-513-1427-3
URN urn:nbn:se:uu:diva-467545 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-467545)

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Ornstein, P., Rickne, J. (2013). When does intimate partner violence continue after separation?. *Violence Against Women*, 19(5):617-633.

II. Ornstein, P., Lyhagen, J. (2016). Asymptotic properties of Spearman's rank correlation for variables with finite support. *PloS one*, 11(1):e0145595.

III. Lyhagen, J., Ornstein, P. (2021). Robust polychoric correlation. *Communications in Statistics-Theory and Methods*, 1-21.

IV. Fogelgren, M., Ornstein, P., Rödin, M., Thoursie, P. S. (2021). Is Supported Employment Effective for Young Adults with Disability Pension? Evidence from a Swedish Randomized Evaluation. *Journal of Human Resources*, 0319-10105R2.

V. Ornstein, P. (2017). *The price of violence: Consequences of violent crime in Sweden*. Working Paper 2017:22, IFAU.

Contents

Introduction

The discipline of statistics is concerned with measurement, from collecting and organizing data to analyzing, interpreting, and presenting it. Because statistics deals with the interpretation of empirical observations, it is integral to all scientific disciplines. The themes in this dissertation span the statistical field, including descriptive statistical inference, detailed questions of measurement and the precise definition of statistical dependence, and statistical models for making causal claims.

In this introduction, I introduce statistics as a tool for empirical analysis, putting the statistical problems discussed in the included articles into the broader context of statistical reasoning. Because of uncertainty in the data, statistical inference is essential to any quantitative analysis. Inference is contingent on the quality and collection of the data. High quality measurement of variables, together with an understanding of how the units (e.g., individuals or firms) were sampled, enable interpretation of the results in a specific sample, potential generalization to a larger population, and sometimes discussion of cause and effect.

This thesis includes five articles in which quantitative methods are applied to data to draw conclusions about the world. Article I introduces descriptive statistical inference in the framework of a decision tool, arguing for the usefulness of statistical models to support expert judgment. Articles II and III focus on statistical measurement. In social science there is often a mismatch between theoretical concepts and their measurement. Articles II and III develop new methods for statistical inference when the real valued theoretical construct is measured on an ordinal scale. In Articles IV and V, the statistical analyses move from models of statistical association to how to handle sampling issues to allow causal interpretation of presented relationships between variables of interest.

Introducing statistical reasoning: describing the world from empirical observations and drawing conclusions beyond that which is directly observed

Article I reports a risk marker study in which the outcome of interest was exposure to assault. The objective was to assess risks against personal safety for *women separated from an abusive partner. Focused on separated couples with a history of domestic violence, the investigation concerned the importance of information about previous controlling behaviors in determining the risk of future stalking or assault.*

Expert judgment is often used in decision making, although statistical models consistently outperform experts in predictive accuracy (Kahneman et al., 2021). Because humans often have preconceived ideas (biases), experts cannot be completely objective in their analysis of data. Furthermore, human judgment is not consistent, and different experts tend to offer different assessments. Statistical models can systematically handle large amounts of data, separate the signals from the noise, and be designed to reduce systemic bias. Another feature of statistical analysis is its ability to quantify uncertainty, letting experts know how certain they can be about the outcome of an assessment. Thus, results of a statistical analysis can inform experts about what to look for and how to compile and interpret gathered information.

Risk marker studies are descriptive studies aimed to identify factors (attributes or characteristics) associated with an outcome of interest. Their results may therefore help determine what information is important to include in police or social services screenings. Although risk factors are not necessarily causal, they can help professionals in assessing the likelihood of future outcomes. The usefulness of risk markers comes from the ease with which they can be observed and the information they provide about the probability of the outcome. A good risk marker combines a substantial increase in information about an outcome of interest with a substantially lower effort than could be achieved using other strategies.

The specific question posed in Article I seeks to determine whether information about the former partner's use of controlling behaviors can improve predictions of post-separation assault: Does the added predictive value of previous controlling behaviors outweigh the increased complexity of including such information in the screening tool? To model this risk assessment using descriptive statistics, denote separation assault as Y and history of controlling behaviors by X. The empirical part of the decision process required measurements (data) of X and Y for a set of individuals drawn from a population. Using the available data, we calculated the associations between X and Y. Modeling post-separation assault as a binary event (occurring or not), the average risk of Y in the population $(P(Y = 1))$ corresponds to its expected value, $E[Y]$. Thus, $E[Y]$ represents the risk of assault to a random individual in the population. The conditional probability of Y given X is the average risk of Y in a subset of units in the population, where the subset is defined by conditioning in the values of X. The analysis of risk of post-separation assault given information about previous controlling behaviors therefore corresponds to the conditional probability of Y given X, $E[Y|X = x]$. For simplicity of notation, let the history of controlling behaviors be binary. The difference between the risk of post-separation assault given controlling behaviors compared with the risk without such history then corresponds to $E[Y|X = 1] – E[Y|X = 0]$.

At the time of screening, future exposure to post-separation assault for a woman selected for screening, will be unknown. We were thus not interested in the individual values in the sample, but simply used them to draw inferences about a (hypothetical) population of all individuals who could be assessed for risk of post-separation assault. The next step in the analysis was to apply statistical inference to understand the value of the information we had and thus the potential of using a history of controlling behaviors to assess the risk of assault for a new individual. While descriptive statistics give information about the relationship between X and Y among the individuals we observe, statistical inference lets us know how much value this information has for assessing risk for any other person in the population.

The specific question posed in Article I was whether we could improve on post-separation assault predictions by including information about the former partner's use of controlling behaviors. We wondered whether the added predictive value of controlling behaviors would outweigh the increased complexity of including such information in a screening tool. We found that a history of control behaviors acts as one of the most important indicators of post-separation risk of stalking and assault, and thus recommended that such information be used in screening.

Measurement and the precise definition of dependence

Articles II and III concern the properties of dependence measures. Article II derives the asymptotic distribution of Spearman's correlation for a particular type of ordinal data. Article III uses the results shown in Article II to construct a new version of another type of correlation measure, polychoric correlation.

Article I on risk assessment focused on the dependency between two variables, the assumption that information about one variable can be used to draw conclusions about the other. The intuitive idea of dependency between two variables is that the larger the value of one, the larger the value of the other, such that knowledge of one is informative about the other. To introduce Articles II and III, it is necessary first to discuss some details about precisely what dependence is and how it is affected by the statistical properties of the data.

While there is a well-defined definition of independence, many definitions of dependency appear in the literature. These slightly different definitions share several common properties, including monotonicity (a positive dependency between two variables also includes a positive dependency between strictly monotone functions of them); symmetry (the dependency between the first and the second variable equals that of the second and the first); and zero dependency given independence (while a dependence measure does not need to capture every relationship between two variables, if they actually are independent, a dependence measure of them must equal zero)[1].

Many measures fulfill these properties and are thereby dependence measures. The most used is Pearson's product moment correlation, also simply called correlation. It measures the strength of the linear relationship between two variables and is thus invariant to change by linear functions of the variables.

Ordinal data

Articles II and III focus on how to measure dependency between variables whose values are ordered, but where the specific numbering of the values does not carry arithmetic meaning. Such variables are called ordinal. The difficulty with ordinal data as opposed to data measured at intervals or on ratio scales is that distance between consecutive differences have no meaning. The distance between Orders 1 and 2 cannot be assumed to be the same as that between Orders 2 and 3; for example, a dinner might be bad, decent, or great, but two decent dinners do not equal one great meal. This fact makes arithmetic operations using ordinal data problematic.

Pearson's correlation, the most common dependence measure in statistics, is calculated using arithmetic operations on variables measured at the interval or ratio scale. It is thus not meaningfully interpretable for ordinal variables. In this case, Spearman's rank correlation can be used instead (Spearman, 1904). Spearman's rank correlation between two variables is equal to Pearson's correlation between the rank values of those two variables, but rather than assessing linear relationships, Spearman's correlation measures the degree to which two variables have a monotonous relationship.

Articles II and III investigate the measurement of dependence on a particular form of ordinal data, the Likert scale. Social science questionnaire items commonly ask respondents to report their agreement with a specific statement on a scale with a finite number of possible answers (e.g., 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree. Such *Likert-scaled* variables are defined by the following three properties: they take a finite number of responses; the responses are ordered, with e.g. lower numbered responses indicating less agreement than higher numbered; and the distances between the actual values of these responses are unknown.

Article II focuses on the asymptotic properties of Spearman's rank correlation between two Likert-scaled variables and derives the asymptotic

[1] See Lehmann (1966) for a probabilistic definition of dependence. A slightly different definition, with importance for Articles II and III, defines *concordance*. See Scarsini (1984), Nelsen (2006), and Nešlehová (2007).

distribution of Spearman's rank correlation for Likert-scaled variables. An asymptotic distribution is important for statistical inference, for example, by allowing for the construction of confidence intervals.

Discrepancies between measurement and theoretical constructs

The human mind has difficulty allocating detailed numerical values to phenomena, even those that could be given such properties (Kahneman et al., 2021). Alignment with a political party or trust in members of one's community is felt as more or less than something other, rather than as an exact point on a scale. This difficulty is true even for phenomena that do exhibit the ratio scale properties of true zero and equal intervals between neighboring points. People are experienced as more or less extroverted, one's own health as better or worse than normal, etc.[2] Collecting data based on such human experience and recollection often involves the Likert-scaled measurement of ratio-scaled theoretical constructs. That is, using discrete and ordinal measurements while theorizing about a continuous phenomenon.

Polychoric correlation is developed to handle this type of discrepancy between measurement and theory. Combining two observed Likert-scaled variables with a distributional assumption allows the calculation of the dependence between the corresponding hypothesized variables of interest. Assuming the hypothesized variables are bivariate normal, polychoric correlation represents their linear correlation. In other words, polychoric correlation is a dependence measure that introduces a parametric assumption to deal with the identification problem that results when observed variables are represented by only a few values, but the variables of interest are continuous. It is based on the idea that the information lost by having only partial information of the value of each observation can be reconstructed by knowledge of the distribution of the underlying variables.[3]

Of several possible estimation strategies for polychoric correlation, maximum likelihood gives the most efficient solution (Olsson, 1979). The assumption of bivariate normality that allows for the calculation of the polychoric correlation, however, may not hold. Other estimation strategies sacrifice some efficiency to gain better robustness against departures from the distributional assumptions (Quiroga, 1992; Roscino & Pollice, 2006). Article III belongs to this literature.

Article III proposes a new estimation strategy for polychoric correlation and gives its asymptotic distribution. A small simulation study showed that the proposed estimation strategy has some advantageous qualities.

---

[2] For example, there is a vast literature on the usefulness of measuring health by Likert-scaled questionnaire items. See e.g. Idler & Benyamini (1997) or McDowell (2006).

[3] In economics, this parametric strategy is used to handle unobserved incomes (See Heckman, 1979).

Modeling questions of cause and effect

In Article I, the focus was on the dependency between two variables. In Articles II and III, the issue of estimating dependency was discussed and problematized in more detail. Often, however, in addition to wanting to know the relationship between variables, we want to know why things happen. Articles IV and V therefore focus on causation. The theme of causal inference introduces a fundamental problem in empirical analysis concerning the difference between the state of the world as it is and the state of the world as it might have been.

Article IV concerns the effect of a specific type of vocational training on labor market participation. A randomized controlled trial was designed in which program participants were randomly assigned to receive the offer of training or not. Randomization makes it possible not only to study the association between vocational training and labor market participation, but also to answer how the training affects employment in the studied population.

In Article V, the effects of non-lethal violent crime on victims are analyzed using observational data to estimate how such victimization affects subsequent health and earnings. To infer causation from observational data, the study attempted to construct independence between potential outcomes and treatment status using a matching design.

A formal language for the hypothetical

Causality has interested philosophers for thousands of years and has been discussed using many different, more or less understandable, definitions. Holland (1986) reviews some of those ideas in the context of what he calls *Rubin's model for causal inference*. Holland argues that Rubin's model adds clarity and concretization to the concept of causality in the sense relevant to statistics, making it useful for statistical inferences with observational data. He distinguishes the idea of an effect of a cause (changes in a response variable attributable to a change in the causative variable), from discussions of possible causes of an effect of interest, and argues that from a statistical standpoint, causality specifically concerns measuring the effects of causes.

Rubin laid out the model for causal inference in 1974 and extended design-based sampling in randomized trials to the use of observational data.[4] The crucial elements of the model were the realization that randomized experiments are embedded in a larger class of allocation or assignment mechanisms and the concept of potential outcomes, which allowed the construction of an outcome variable for each cause or treatment.

[4] Hypothetical outcomes in the case of the randomized experiment were first formalized by Neyman in 1923 (Neyman et. al., 1990). Later, but more easily accessible works by Neyman and by Fisher, however, were more important to generalizing the idea of hypothetical outcomes to observational studies. (See Rubin, 1990).

The causal question in Article IV, the effect of training on employment, addressed the difference between the employment status of individual participants and their expected employment status had they not taken part in training. We thus have two outcomes, employment following training, *Y(1),* and employment in case of no training, *Y(0)*. In asking about the effect of training on employment, we wonder about the difference between these two outcomes. Looking specifically at individual *i*, denote this difference Δ_i:

$$\Delta_i = Y_i(1) - Y_i(0)$$

The difference between the potential outcomes equals the causal effect of vocational training on employment. This explicit and formal definition of the causal effect, Δ_i, makes it easy to see whether a scientific question is or is not about causality. We can also clearly see why the individual causal effect is fundamentally unobservable. Even when we observe the employment status for each individual in a sample, half of the observations in the potential outcomes *Y(1)* and *Y(0)* are missing. For each individual *i*, either $Y_i(1)$ or $Y_i(0)$ can be observed, but never both.

Moving toward the observable

Denote the observed training status by *D*. For ease of representation, we assume no interference between units, so the outcomes of one individual is affected only by that individual's treatment assignment. In addition, we want to make sure that all available ways to take part in training have a representation in D. In our case, since D is binary, that implies assuming training either takes place or not. These assumptions about the treatment, that it is fully represented with no unknown treatment options, and that it affects the treated unit only, ascertains that the value of the outcome $Y_i(D)$ is determined. It is called the *stable unit treatment assumption, SUTVA* (Cox, 1958; Rubin 1980).[5]

Under SUTVA, the observed outcome can be written as a joint function of the observed training status and the potential outcomes. This formalizes the relationship between the outcomes of interest, *Y(1)* and *Y(0)*, and the observed outcome, *Y*. For any individual *i*, we could observe employment after training if they participated ($D=1$), or employment after no training if they did not, ($D=0$):

$$Y_i = Y_i(1)D_i + (1-D_i)Y_i(0)$$

Thus, the difficulty of causal inference can be understood as a missing data problem. Categorizing the fundamental and philosophical problem of causal

[5] See also Rubin (1986) and Rubin (2010) for clear and accessible explanations of SUTVA.

inference as a missing data problem means that we take one more step from the philosophically hypothetical toward a model that allows for measurement.

The expected value $E[Y]$ is defined as the average value of Y over all individuals (units) in a given population. The average causal effect of vocational training on employment, Δ, can be defined as the difference between the expected value of $Y(1)$ ($E[Y(1)]$), and the expected value of $Y(0)$ ($E[Y(0)]$):

$$\Delta = E[Y(1)] - E[Y(0)]$$

Because of the fundamental problem of causal inference, unit level effects cannot be observed so Δ_i is always unobservable. However, $Y(1)$ can be observed from some units and $Y(0)$ from others, implying the possibility of deriving knowledge about Δ. The question is what we are willing to assume about the pattern of observation and the missing values of $Y(1)$ and $Y(0)$.

Randomization

Using missing data terminology, if observations are missing completely at random, we can safely disregard them. Randomization between training and no training means that this will be the case. We can treat the missing observations as missing completely at random, which implies that the missing observations can be omitted without affecting the expected value. Given randomization, the potential outcomes are independent of training, which means that the expected value of the observed outcomes equal the expected value of the potential outcomes, that is $E[Y|D=1] = E[Y(1)]$ and $E[Y|D = 0] = E[Y(0)]$. Under randomization, what can be observed equals what we need, and Δ can therefore be expressed using only observable terms:

$$\Delta = E[Y(1)|D = 1] - E[Y(0)|D = 0]$$

In Article IV, this solution was chosen to estimate the average effect Δ of the cause training on the outcome employment. As treatment allocation was randomized, the empirical analysis was straightforward, and statistical association could be interpreted using the framework of causal inference.

Independence through ex-post design

In the section above, randomization is shown as a solution to the problem of causality. Randomization works because it introduces independence between the potential outcomes and exposure status. Therefore, exposed and non-exposed individuals differ on average only in terms of their treatment status prior to exposure, so their outcomes can be interpreted as causal effects of treatment or non-treatment.

It is not always possible to conduct a randomized experiment, and even when it is possible, randomization may introduce new problems. Fortunately, it is not randomization, but the independence it creates, that solves the missing data problem. Experiments are just as valid if units are divided before exposure by another logic into comparable groups (see, e.g., Deaton & Cartwright, 2018). For example, an agricultural experiment might divide a plot of land into a chessboard pattern, with what would be white squares treated and the others kept as controls. Such a pattern is not random, but it is orthogonal to the potential outcomes.

Orthogonality between exposure and potential outcomes holds for any type of situation that allows comparisons between individuals who are truly similar on every pre-exposure characteristic, as truly similar individuals have the same distributions of *Y(0)* and *Y(1)*. This implies the following: Assume we know everything that is simultaneously relevant to both the assignment to treatment and the potential outcomes and that this information is captured by *X*. Then, if we condition on *X*, the expected values of the potential outcomes *Y(1)* and *Y(0)* will be independent of treatment status, *D*, for any given value of *X*, which in turn will allow for identification of *Δ*.

$$E[Y(1)|X] - E[Y(0)|X] = E[Y(1)|D = 1, X] - E[Y(0)|D = 0, X]$$

Here the expectations are over the distribution of *X*. For a given value of *X*, which contains all characteristics that affect the values of *Y(0)* and *Y(1)*, we have the independence needed to condition on *D* while not affecting the expected values of *Y(0)* and *Y(1)*. This allows us to identify the causal effect of exposure for a given X, as the difference on the right-hand side is observable.

The point of this section is to show that causal effects can be identified through post-exposure designs. However, to achieve this empirically, we need to take a step away from the directly observable to analyze results that hold only under strong assumptions. As we can only condition on what we can observe in data in the estimation, we need to assume that all relevant characteristics that simultaneously affect *D* and either *Y(0)* or *Y(1)* are observed.

Implementations of holding constant for factors potentially affecting the outcome value are most well-known in the form of linear regression analysis. If the linear regression model is correct, the outcome is compared between individuals who are similar on those factors. The first problem is that the design (i.e., the linear model) may not be a correct representation of the data. In addition, specifying the model after the data has been collected risks compromising the design and hence any inferences drawn from the analysis. The researcher cannot understand the sampling conditions without simultaneously learning about the existing associations and which design choices will produce the strongest results. A better ex-post design strategy is to place design distinctly ahead of analysis (Gelman & Loken, 2013; Rosenbaum, 2010).[6] One strategy that allows for such a separation of design and analysis is *matching*.

Matching is the idea of looking for statistical *twins* who differ in exposure status but are indistinguishable on pre-exposure characteristics. The more similar they are, and the richer the set of observed variables that exist to examine their pre-exposure likeness, the more convincing the matching strategy, but the greater the difficulty of finding individuals who match the strict criteria. However, there is no need to find actual twins of this type, one treated and one non-treated. It is enough for the variable X to be – on average – similarly distributed over the treated and non-treated groups. In fact, it is enough for treated and non-treated individuals to have the same risk of exposure given X. This is the propensity score theorem of Rosenbaum and Rubin (1983) that states that the only characteristic that needs to be conditioned on is the exposure risk. In this type of analysis, the exposure risk is called the *propensity score*, denoted as $p(X)$. This measure summarizes the relevant information of the observed covariates X into a single probability. For exposed and non-exposed individuals with the same propensity score, the distributions of the covariates X are the same (Rosenbaum & Rubin, 1983). The propensity score can be used to select comparison groups of exposed and unexposed individuals that are equal in terms of observable characteristics.

Not an innocuous assumption

The last section has introduced a strategy of causal analysis conducted with only observational data. This strategy assumes we can observe all factors that simultaneously matter to $Y(1)$, $Y(0)$, and D. The main problem with this assumption is that it cannot be tested using data. For this reason, we need to evaluate the design to learn how sensitive the results would be if the assumption were to fail.

Ideally we would like to ensure that the distributions of $Y(1)$ and of $Y(0)$ were the same irrespective of exposure status. As we have a missing data problem, with $Y(0)$ unobservable for those exposed and $Y(1)$ unobservable for those unexposed, this cannot be done. To evaluate the design, however, we can compare the distribution of X between the exposed and non-exposed groups. If X were observed in a well-designed randomized controlled trial, we would expect balance between treated and non-treated. The same thing is expected in a well-designed observational study. Examining observed variables that were not part of the construction of the propensity score is one way to evaluate the trustworthiness of a causal claim. If outcomes can be observed both before and after exposure, and are not part of the propensity score, then

[6] For an example of design decided at a stage after data collection, but before analysis, see Fowler et al. (2017).

placebo testing is another formulation of the same idea: before treatment, outcomes of those treated should not differ from outcomes of controls.

After evaluating the design with satisfactory results, the assumption of conditional independence of the outcome distributions is strengthened, but individuals may not actually be as comparable as they appear on the observed covariates. Another strategy to support a causal interpretation of results is to examine their robustness to deviations against the assumption of conditional independence. How sensitive are the conclusions to deviations against the assumption that we have controlled for all essential variables affecting both treatment and the potential outcomes? The first formal implementation of this type of sensitivity analysis was done by Cornfield et al. (1959), examining the effect of smoking on lung cancer. Their design was further developed by Rosenbaum (1991; 2002), who formulated a general strategy to calculate the robustness of results for unmeasured characteristics that simultaneously affect both exposure status and outcomes. The strategy in this type of sensitivity analysis is to manipulate the estimated odds of exposure (in Article V, the likelihood of assault) to analyze how the effects vary, by asking how an unmeasured covariate would have to be distributed over cases and controls to alter the conclusions of the study. The analysis of how changes in the odds of exposure affect inferences about the impact of exposure is thus used as a measure of sensitivity to hidden, or unobserved, bias. It gives information about the maximum magnitude of bias that could be present without altering the conclusion of a study, denoted by the term *Rosenbaum bonds*.

Article V followed these steps of causal inference from observational data. The propensity score was defined as the probability of assault. During the design phase of the analysis in which the matching estimator was selected, assaulted and non-assaulted individuals with the same propensity score were compared on observable characteristics over the 4 years preceding the year of the assault. Assaulted and non-assaulted statistical twins were then followed up to 8 years past the assault year. Those assaulted were found to have worse outcomes in both health and labor market status. Finally, the robustness of the findings against unobserved systematic differences in assault risk between assaulted and non-assaulted were examined and Rosenbaum bonds calculated, which supported the reported conclusions.

Summary of articles

Article I

In this article, Johanna Rickne and I suggest that a history of controlling behaviors by a former abusive partner is a key predictor of post-separation stalking and assault. Our test was carried out in a cross-sectional sample of 714 divorced or separated Swedish couples surveyed in 2001. The data are from a subsample of the *Swedish National Violence Against Women Survey*, which provides lifetime experiences of violence in a nationally representative sample of 6,926 women. From this survey, we selected a subsample of heterosexual couples in whom the woman was subjected to one or more acts of violence during cohabitation. Thus, we limited the sample to the group of women most at risk of spousal assault after separation, allowing us to focus on the potentially important heterogeneity within this group.

Controlling behaviors such as social isolation, economic constraint, and threats of violence have been found to predict the severity and frequency of violence in ongoing relationships with quite good accuracy. From a rational choice perspective, control is key and domestic violence is understood as a means of controlling one's partner. In this article, we argued that controlling behaviors form part of an overall strategy of sexual coercion, wherein the important heterogeneity of control behaviors can be captured by a one-dimensional measure. We operationalized control using an index of 10 specific controlling behaviors. More than half of the respondents had experienced jealousy, been called disparaging names, and been made to feel inferior and useless. Less common control behaviors included restrictions on the respondents' social and working lives. The use of a one-dimensional measure was supported by a principal component analysis indicating that the index retained 80% of the variance in control.

The empirical results show that men who use controlling behaviors with their partners during a relationship are substantially more likely to commit acts of stalking and violence after separation. In the case of stalking, we found that the likelihood increases linearly with the number of control behaviors exhibited during the relationship. For violence, the relationship seems nonlinear. Low levels of control were found to be unrelated to post-separation assault, but the risk increased for more controlling men with each additional identified behavior. Our basic measures of control behaviors explained 18% of the variance in stalking and 8% of the variance in assaults. The analysis indicates that control behaviors are quantitatively more important predictors of post-separation stalking and assault than several other well-established risk markers such as low income, alcohol abuse, joint children, or violent experiences in childhood.

This article contributes to the literature on the nature of domestic violence. Identifying the important role of control in predicting violent behavior accords with the understanding of control as a coercive strategy in feminist theory and evolutionary psychology and is in line with a rational choice perspective on human behavior. The results also support the usefulness of variables describing control to explain the heterogeneity of the severity and frequency of violent outcomes in relationships. In this article, we extended previous insights from studies of ongoing relationships to post-separation stalking and violence. The article also has implications for preventive work, informing the risk analyses that are part of police, welfare, and judiciary systems. The finding that controlling behaviors are quantitatively more important predictors than several other well-established risk markers highlights their potential to improve risk assessments and therefore help to prevent assault and stalking.

Article II

In Article II, Johan Lyhagen and I derive the asymptotic distribution of Spearman's rank correlation for ordinal variables that take a finite number of values.

Spearman's rank correlation measures agreement between the rankings of two variables, the degree to which they have a monotonous relationship. It equals Pearson's correlation between the rank values of two variables. When the variables are independent and no observation has ties, it follows a normal distribution. However, it is often used specifically for Likert-type variables with finite support. In those cases, the number of ties of each observation grows as the sample size increases. For such variables, distribution theory for Spearman's correlation under dependence is very complicated and precludes both noncentral tests of correlation and the construction of confidence intervals.

The relationship between two variables with finite support can be represented in a contingency table, as has been applied to define Spearman's correlation using copulas. Using that definition and the delta theorem, we showed that Spearman's correlation is consistent and asymptotically normal, and we derived the asymptotic variance. A small simulation study on rejection rates and power indicated that the derived asymptotic variance performs on par with bootstrapping for moderate to large sample sizes.

The practical contribution of the article is to make inference on Spearman's rank correlation straightforward, an improvement specifically useful when constructing confidence intervals or testing hypotheses other than independence. It also increases the potential uses of Spearman's rank correlation in constructing other estimators.

Article III

To handle Likert-type variables, tetrachoric correlation has been extended to *polychoric correlation*, attained through several possible estimation strategies with different properties.

In Article III, Johan Lyhagen and I used the results from Article II to propose a new estimator for the polychoric correlation. We developed an estimation strategy based on minimizing the squared distance between the estimated Spearman's rank correlation and the Spearman's rank correlation as a function of underlying, continuous variables. This rank-based estimator was shown to be consistent and asymptotically normal.

Our choice of this estimation strategy was motivated by the hope that the polychoric correlation might inherit some of the robustness of Spearman's rank correlation. In a simulation study, we compared the robustness properties of the new rank-based estimator with those of the common maximum likelihood estimator. The rank-based estimator outperformed maximum likelihood for most combinations of skewness and category probabilities, with only small efficiency losses when maximum likelihood performed well.

Article IV

In Paper IV, Mattias Fogelgren, Magnus Rödin, Peter Skogman Thoursie, and I report the results of a large-scale nationwide randomized experiment we conducted in collaboration between the Swedish Insurance Agency and the Swedish Public Employment Service. We evaluated whether a supported employment rehabilitation intervention for young adults on disability pension could improve their labor market opportunities better than regular vocational rehabilitation.

Supported employment aims to place the individual at a specific workplace early in their rehabilitation, with a caseworker closely attached to the intervention. In this intervention the caseworker acts as a back-up for the supported individual during training to reduce the employer's risk when hiring an individual with unclear productivity. The control group comprised people in regular vocational rehabilitation, which is less staff intensive and includes in-house work preparation and training, but not support at the workplace. A second control group included participants in a third intervention, case management. Case management requires roughly the same resources as supported employment but takes the broader perspective of supporting individuals in learning how to navigate their lives as opposed to focusing on their work.

A total of 1,062 individuals were randomly assigned to the interventions. The empirical analysis followed an intention-to-treat design, with participants grouped according to their randomization outcome, irrespective of their actual intervention take-up. The main results showed that 18 months after the start of the project, participants offered supported employment had reached employment rates approximately 10 percentage points higher than participants who were offered regular rehabilitation only.

The target population in this study had a wider range of disabilities than usual in research on supported employment, which has traditionally been evaluated in people with severe mental diagnoses such as psychosis. The study thus contributes to the literature by significantly broadening the scope of supported employment as a vocational training strategy, indicating its potential to be successful for the whole population of young jobseekers with disabilities.

Article V

The aim of this study was to fill the gap in the literature on the consequences of violence on victimized people by identifying, estimating, and monetizing the long-term consequences of interpersonal violence on victims.

Previous studies on the relationship between exposure to violence and its implications for victims have either (1) not attempted to establish a causal relationship or (2) used very limited selection criteria, focusing on specific assault characteristics, outcomes, or target populations. Several strategies used to gain at least some basis for policy analysis have completely disregarded the consequences of violence on the victims. It has therefore been impossible to make evidence-based cost estimates of the burden of violent crime on society.

Using rich administrative population data for Sweden from 1994 to 2010, I matched individuals who went to a hospital in the years 1998 to 2002 with injuries from assault with individuals who were statistically indistinguishable from them. The basic idea of matching is to twin individuals meeting selection criteria with others who are similar in all relevant aspects other than the main criterion of interest. In this study, relevant similarity included characteristics that influence the risks both of exposure to violence and of adverse labor market or health outcomes. Therefore, to ensure similarity between assaulted and control individuals, they were matched on factors related to health, family situation, productivity, and demographics in the 4 years prior to the hospital visit.

The results suggest that violent crime has large and persistent effects on mortality, suicide, earnings, work status, disposable income, and number of days on sick leave. Specifically, assaults leading to a hospital visit are estimated to cost 1.4 million SEK per victimized woman and 1.5 million SEK per victimized man; more than 80 percent of these costs are due to excess mortality. Estimates of socioeconomic outcomes show some robustness against selection on unobserved characteristics. Estimates on mortality and suicides are very robust.

Comparing the results of this study with previous reports on the costs of assault shows that indirect costs specific to victims of violence greatly surpass the direct costs incurred by health care institutions or the legal system. The results thus indicate that costs carried by victims of violence form a key component for future cost calculations of violent crime. This study has implications for policy decisions regarding the expected cost-effectiveness of future reforms aimed to increase public and domestic safety.

Acknowledgement

The completion of this PhD-thesis has been a challenging and drawn-out process, but mainly and importantly, it has been a process of learning. During these years, so many people have, in one way or another, contributed. This work would not have been possible without that guidance and help. I would like to express my gratitude.

My advisor Johan Lyhagen has shown encouragement and support, presented suggestions for research ideas, listened to and given space to mine, as well as collaborated with me on paper II and III. Thank you, Johan, I could not have done this without your positive attitude. My assistant advisor, Per Johansson, helped me visit the Department of economics at Columbia University, helped me get access to data, and has been a careful and constructive reader and commentor of paper V as well as the thesis' introduction. Being able to work on paper V was very important for my motivation as a PhD student, and your generosity, knowledge and curiosity, Per, were crucial in making that happen. I am grateful as well to the Economics department at Columbia for letting me visit for a year, to Tom Hedelius' foundation for making that possible, and in particular to Lena Edlund who invited me and offered her time and advice. Thank you, Lena, especially for your inspiring lectures and for your emphasis on not to settle, but to go for clear, intriguing questions and high-quality data. Before starting my PhD, while still thinking about what to study, and why, I found a quantitative report by Eva Lundgren's otherwise mostly qualitative research group. It was such an inspiring read, and I wanted to be part of that intellectual context. I am so grateful to you Eva for generously inviting me into your group, and for sharing data and insights.

My colleagues at the Statistics department at Uppsala university, thank you for making me feel welcome. Special thanks to Katrin Kraus, for a friendly introduction, to Hao Luo, for the friendly office atmosphere, and to Lisbeth Hansson, Dag Sörbom and Thommy Perlinger, for organizing a teaching experience that I enjoyed and learned from.

I would also like to thank my coauthors. Thank you for exploring together with me! Especially I'd like to thank Johanna Rickne, who collaborated with me on paper I. I have enjoyed your thoughts and company, admired and learned from your ability to focus and to wrap-up and finalize projects, and benefited greatly from your encouragement to apply for grants and to believe in my projects. This PhD would have been less fun, and involved less learning, without your support. Peje Bengtsson, I could not have been luckier than to have had you at my first non-academic workplace. Our academic collaboration was in the typical style of your mentorship, where you had pawed the way for an interesting project, and then opened the space for junior colleagues such as me. Peter Thoursie and Magnus Rödin, who have worked with me on paper IV, thank you for offering a kind, encouraging and intellectually curious work environment. I am so enjoying our current project and looking forward to several follow-up studies over the coming years.

I would also like to thank my ”near-coauthors”, and “coauthors-to-be”. To Patric Tirmén and Maria Cheung, it has been a rewarding experience to work close together with you, but also later to follow your respective work and share ideas from a distance. Anna Lexelius and Johanna Thagemark, and thereafter Josefine Palmqvist Schultz, working together with you have been among my most rewarding experiences. Our collaborations have opened me to a different and broader view of my work, my abilities, and my passions, than what I could have found on my own, and that perspective has contributed to my finishing this thesis. Thank you. Åsa Eldén, I think our interdisciplinary collaboration worked so well due to your open-mindedness. I hope we will find the time to continue. Eva Ranehill, you steer our project forward with a structure I can only admire, and you combine it with humility and kindness. It is a true pleasure to work with you. Erik Jönsson, Petter Helgesson, and Ulfhild Westin – you each have skills that I do not have, so that having you on my team, I can propose ideas that would otherwise just be irrelevant – I am so grateful to be allowed to give back some of what I managed to learn from Peje, in the joyful and productive context of great colleagues like you.

Finally, thanks to family and friends who have supported me throughout this project. Special thanks to Elin and Fredrik for stubborn help in finalizing the dissertation: for access to a writing cabin, whip and carrots, deadlines, and rest. Thanks! Thank you also Anna Ardin, Sara Fogelberg, Kajsa Hallberg and Anna Lindgren, among others, for your respective contributions throughout the years. And lastly, thank you Lo, and you Sasha, for making life, and all its content, more meaningful.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-455.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, *22*(1), 173-203.

Cox, D. R. (1958). *Planning of experiments*. Wiley.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2-21.

Fowler, P., de Luna, X., Johansson, P., Ornstein, P., Bill, S., & Bengtsson, P. (2017). Study protocol for the evaluation of a vocational rehabilitation. *Observational Studies*, *3*, 1-27.

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153-61.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945-960.

Holland, P. W., & Rubin, D. B. (1983). On Lord's paradox. In H. Wainer & S. Messick (eds.), *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord*. Routledge, pp. 3-25.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*(1), 21-37.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.

Kendall, M. (1948). *Rank correlation methods*. Griffin.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, *37*(5), 1137-1153.

McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires*. Oxford University Press.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, *98*(3):544-567.

Neyman, J., Dabrowska, D. M., & Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, *5*(4), 465-472.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, 44(4):443-460.

Quiroga, A. (1992). *Studies of the polychoric correlation and other correlation measures for ordinal variables* [doctoral thesis]. Uppsala University.

Roscino, A. & Pollice, A. (2006). A generalization of the polychoric correlation coefficient. In *Data analysis, classification and the forward search: Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, University of Parma, June 6-8, 2005*. Springer, pp. 135-142.

Rosenbaum, P. R. (1991). Sensitivity analysis for matched case-control studies. *Biometrics*, 87-100.

Rosenbaum, P. R. (2002). Sensitivity to hidden bias. In *Observational Studies*. Springer Series in Statistics, Springer, pp. 105-170.

Rosenbaum, P. R. (2010). *Design of observational studies* (Vol. 10). Springer.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test. Comment. *Journal of the American Statistical Association*, *75*(371), 591-593.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961-962.

Rubin, D. B. (1990). [On the application of probability theory to agricultural experiments. Essay on principles. Section 9.] Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, *5*(4), 472-480.

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010).

Scarsini, M. (1984). On measures of concordance. *Stochastica*, *8*(3), 201-218.

Spearman C.E, 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15: 72–101.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 195

Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-467545

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022