Dear Commissioners, I note with interest the concerns raised in your draft report about the restrictions on data analysis that occur due to privacy issues. In light of these concerns, you may be interested in the attached draft paper. It proposes an estimator which can be applied almost anywhere that a logit or probit estimator can be applied, but does not require individual-level data. Instead, the estimator works with data which has been aggregated by strata, where the strata are constructed as the intersection of the control variables. If you are interested in further explanation or example code I would be happy to help. We plan on submitting the attached working paper to a leading econometrics journal within the next few months. Sincerely, Emma Aisbett

7 October 2016

Estimating Treatment Effects from Counts of Binary Outcomes: a Conditional Likelihood Estimator of Relative Risk*

Christopher Aisbett[†], Emma Aisbett[‡] and Ralf Steinhauser[§]

October 5, 2016

Abstract

This paper introduces the conditional likelihood estimator of relative risk (CLERR). The CLERR estimates the relative risk of an outcome analogously to the way the conditional logit estimates an odds ratio. Aside from the fact that relative risk is often the preferred measure of association, the CLERR has superior statistical properties, including both unbiasedness and efficiency in small and large samples. The CLERR can be thought of as an exact matching approach which allows estimation of treatment effects for binary outcomes, without the need for structural assumptions about the influence of confounding variables. We apply the CLERR to World Bank's Enterprise Survey data and show that firms with female owners are significantly more likely to export than similar firms with only male owners.

Keywords: conditional likelihood estimator, relative risk, binary outcomes, treatment effects, matching, female owners, exporting

*Preliminarly Draft. In loving memory of Christopher Aisbett, who came up with the idea for this estimator and identified many of its favourable statistical properties.

[†]Laeta Pty. Ltd.

[‡]Faculty of Business and Economics, University of Hamburg, Welckerstr. 8, R 2.20 D-20354 Hamburg. Tel.: +49 40 42838 8636. Email:Emma.Aisbett@wiso.uni-hamburg.de

[§]Faculty of Business and Economics, University of Hamburg

1 Introduction

Estimation of the causal impact of a binary variable (i.e. treatment effect) is the objective of a substantial proportion of statistical analyses across both the social and biological sciences. 'Treatment' variables, of course, are not restricted to medical or policy interventions, but may be any categorical variable which influences the outcome of interest. In the social sciences, treatment variables for individuals include variables such as gender, race, citizenship, and political affiliation. For firms, examples of treatment variables are foreign ownership, CEO gender, or incorporation in a particular state. The current paper introduces an efficient and unbiased estimator of treatment effects on a binary outcome variable — measured as relative risk. Examples of applications to which the proposed estimator is ideally suited include estimating the extent to which certain 'treatments' affect the probability of outcomes such as giving to charities, evading taxes, having a car accident, becoming unemployed, supporting action on climate change, exporting, or gaining certification.[1]

In economics and the social sciences at large, treatment effects for binary outcomes are typically estimated using either generalized linear models (especially Poisson, logit and probit), or specialized treatment-effects models (including propensity-score matching and inverse-probability weighted approaches). Most of the common treatment-effect methods themselves involve some form of generalized linear estimation approach for either the selection-into-treatment or the outcome model (and sometimes both). These modeling approaches are necessary when exact (or coarsened exact) matching is precluded by a lack of sufficient observations within matched groups. The price paid for their structural assumptions is the potential for specification bias and an increased scope for researchers to choose specifications which suit their priors.[2]

As data collection becomes ever cheaper and the volume of data available for analysis ever greater, it is timely to consider approaches to estimating treatment effects which exploit exact matching of treated and control observations within a strata[3], and thus reduce the need for structural assumptions.

[1] Classic examples from the health, medical and biological sciences include mortality, disease incidence, and possession of certain traits.

[2] See Ho et al. (2007) for a thorough discussion of these issues.

[3] In other literatures and contexts these 'matched groups' are often referred to as 'strata' or 'conditioning groups'. We will use all three labels interchangeably.

The conditional logit estimator has long been the stalwart of such analyses for binary dependent variables.[4] The current paper proposes an attractive alternative — the conditional likelihood estimator of relative risk.

While conditional likelihood estimator of relative risk (CLERR) is conceptually very similar to the conditional logit estimator, it has a number of advantages of both statistical and practical nature. To begin with, the conditional logit suffers from small-sample bias as well as asymptotic inefficiency.[5] Bias in the conditional logit estimates are particularly a problem when the number of control variables (and thus conditioning groups) is large relative to the number of observations. That is, attempts to reduce omitted variable bias may increase estimator bias when the conditional logit is used. The CLERR, on the other hand, does not suffer from such problems. It is unbiased and efficient in both small and large samples. Specifically of relevance for small samples, the CLERR's estimating equation is the unique unconditionally optimal one. The CLERR's large-sample properties include asymptotic normality, consistency, and achievement of the Cramer-Rao lower bound for asymptotic efficiency.[6]

The most obvious difference between the CLERR and the conditional logit estimator is that the CLERR estimates relative risk, while logit models estimate odds ratios.[7] Odd-ratios are unavoidable in case-control studies, but in cohort data (typical in the social sciences), they have little to recommend

[4]See for example Breslow and Day (1980); Kupper (2005).

[5]For discussion and examples of sometimes substantial small-sample bias of the conditional logit estimator, see Greenland (2000); Greenland, Schwartzbaum, and Finkle (2000). Andersen (1970, p. 299) shows that the conditional maximum likelihood estimator of an odds ratio does not achieve the Cramer-Rao lower bound for asymptotic variance. This contrasts with authors such as McFadden (1974) who say the conditional logit estimator is asymptotically efficient. The difference between the two conclusions about the conditional logit's efficiency arises because Andersen considers efficiency relative to the full information set, whereas McFadden considers it only relative to the information remaining after conditioning out the incidental parameters. Andersen's result shows that some information is lost during conditioning because the sufficient statistic is not ancillary to the log odds ratio (which is what the conditional logit model estimates). See also Breslow (1981, p.77) on this point. In contrast, the sufficient statistic for the strata is in fact ancillary (according to Andersen's definition) for the relative risk (which is what the CLERR estimates). Thus no information is lost in conditioning using the CLERR.

[6]We prove these properties in Section 2.

[7]'Relative risk' is alternatively known as 'relative incidence rate', 'relative prevalence' or 'prevalence ratio' depending on the context and the literature. For our purposes they are synonymous.

them over relative risk (Greenland and Robins, 1985).[8]

One of the key advantages of relative risk is its intuitive appeal — an important feature when communicating research findings. Relative risk is simply defined such that the probability of a positive outcome for a member of the treatment group equals the relative risk times the probability of a positive outcome in the control group: $Pr(outcome|treatment) = p.r$, where p is the baseline risk and r is the relative risk. For odds ratios, the probability of a positive outcome for a member of the treatment group is difficult to even explain without resort to mathematics: $Pr(outcome|treatment) = \frac{p.OR}{(1-p)+p.OR}$, where p is the baseline risk and OR is the odds ratio.

From the perspective of the research community, perhaps a more important advantage of relative risk compared to the odds ratio is that relative risk has the property of collapsability. In practice, collapsability means that the true relative risk will not change as the definition (and number) of conditioning groups change. Accordingly, the estimated relative risk cannot be distorted by controlling for excess groups through introduction of irrelevant controls to the conditioning-group definition. The same is not true for odds ratios.[9]

Of course, when the baseline probability of a positive outcome is sufficiently small, the estimated odds ratio will be very close to the relative risk.[10] Even in these cases, however, the CLERR has advantages over the conditional logit estimator. Aside from the superior statistical properties mentioned earlier, the key advantage of the CLERR is that it can be applied not only to individual-level data, but also to data which has been aggregated at the strata level. In many cases, statistical agencies and data collection firms have individual-level data, but access to such for researchers is restricted due to privacy concerns or is prohibitively expensive. The CLERR can use count data, which has been aggregated at the strata level, to examine the

[8]Papers discussing the relative merits of alternative means of estimating relative risk and discussing the limitations of odds-ratio estimators such as the conditional logit include Barros and Hirakata (2003); McNutt, Wu, Xue, and Hafner (2003); Coutinho, Scazufca, and Menezes (2008); Cummings (2009); Lee, Tan, and Chia (2009); Diaz-Quijano (2012). Others, such as Breslow and Day (1980) assume that relative risk is the preferred measure, but rely on the 'rare disease' assumption to justify equating the two.

[9]For example, if the researcher increased the number of conditioning groups by controlling for marital status, the true odds ratio would change even if marital status had no actual impact on the probability of a positive outcome. In this same situation, the true relative risk would not change.

[10]Synonyms for 'positive outcomes' in other literatures include 'successes' and 'events'.

determinants of binary outcomes at the individual level.

The typical approach in the social sciences literature to estimation of relative risk from count data involves application of a Poisson or related model. Aside from the potential biases mentioned earlier for such generalized linear models, Poisson models have an additional bias if the risk of a positive outcome is not small. The source of this bias is that the (true) Binomial distribution only converges towards the Poisson distribution as the number of trials goes to infinity while the product of the number of trials and the probability of a positive outcome remains fixed. Thus the Poisson distribution is best applied to systems with a large number of possible events, each of which is rare. The CLERR does not have this limitation, as it is derived directly from the underlying Binomial distribution.

Another estimator sometimes used in the literature for estimating relative risk is the "fixed effects" Poisson estimator of Hausman, Hall, and Griliches (1984). This estimator is, however, only applicable for conditioning out strata effects in the special case of matched-pair (1:1 matched) case-control data. In this special case, both our estimator and that of Hausman et al. (1984) reduce to the simple ratio of the unweighted sum (over all strata) of the events in the treatment group divided by the equivalent sum for the control group. In other words, the unconditional ratio of mean number of counts is the maximum likelihood estimator of true ratio when the dataset is comprised only of matched case-control pairs.[11]

A natural question is what happens if you use a conditional Poisson model which is flexible enough to allow for varying numbers of members of the treatment and control groups in each strata? The answer is that you get exactly the CLERR. The conditional likelihood equation is the same whether the marginal distribution of the counts is assumed to be Poisson or Binomial.[12] On this basis, the CLERR can be considered a conditional Poisson estimator, or a conditional Binomial estimator, of relative risk. Our proofs of the CLERR's properties, however, assume the underlying data are binomially

[11]See for example Section 9 of Andersen (1970) for proof of this result for the Poisson case.

[12]In fact, in the natural science literature on relative risk, an estimating equation equivalent (but not identical) to ours is mentioned as the solution to the Poisson conditional relative risk estimating equation by Gart (1978). Gart does not, however, investigate the estimator's properties. Breslow (1984); Greenland and Robins (1985) mention Gart's estimator, but pay it little attention, apparently because of the lack of an explicit solution to the estimating equation.

distributed.

Having introduced the estimator and proved its properties in the first half of the paper, in the second half we demonstrate the application of the CLERR using data from the World Bank's Business Environment Surveys. The standardized core questionaire dataset includes responses from over 100,000 firms across 135 countries from 2006-2014. We use the data to examine the relative propensity of firms with at least one female owner to engage in exporting. We examine the relative risk of exporting for firms with female owners, controlling for confounding variables using several different strata definitions. The most conservative of these constructs strata from the intersection of the sector (e.g. textile and clothing), legal status (e.g. sole proprietorship), size (quintiles), and location within country. This specification results in nearly 30,000 strata populated with at least one observation (and many more if empty strata are counted).

We find female-owner firms are consistently at least as likely to engage in exporting as all-male owned firms. Furthermore, when we focus on firms where owners have more control — by excluding subsidiaries and firms with government ownership — we find robust evidence that female-owner firms are statistically significantly more likely to export. CLERR estimates suggest that firms with female owners are $1.09 - 1.13$ times as likely to engage in exporting than otherwise similar firms.[13] Among sole proprietorships and general partnerships, the relative risk of exporting is almost 30% higher for female-owner firms than all-male firms.[14] We posit that the higher propensity to export among firms with female owners is due to women being generally more risk averse (Eckel and Grossman, 2008), and exporting being a means of diversification which increases the stability of firm profits (Hirsch and Lev, 1971; Miller and Pras, 1980).

For comparison, we also conduct the analyses using the conditional logit estimator. Based on the results from the conditional logit estimator, one could conclude that firms with female owners are up to 1.46 times more likely to engage in exporting.[15] It is difficult to judge the extent to which

[13]Based on estimates in Table 3.

[14]The relative risk estimates for general partnerships and sole proprietorships were 1.29 and 1.28, respectively.

[15]This conclusion would be reached if one followed the common practice of making the "rare disease" assumption and interpreting the odds ratios in Table 4 as if they were relative risks. Given that it is not possible to convert odds ratios from the conditional logit estimator, there is no other way one could make any readily interpretable statement

the conditional logit results differ from those of the CLERR due to biased estimates of the odds ratio, or due to genuine differences between the odds ratio and relative risk.[16]

In terms of mechanical performance, the CLERR produced estimates at least as quickly as the conditional logit. Both the CLERR and the conditional logit always converged without problem. In contrast, poor or extremely slow convergence characterised almost every other estimation approach we tried.[17]

2 The Conditional Likelihood Estimator of Relative Risk (CLERR)

We are interested in estimating a relative risk (or relative incidence rate, or relative prevalence).[18] Let r be the true relative risk of an event (e.g. death from heart disease) occurring in the 'treated' versus control group (e.g. people with and without a parent who died from heart disease).[19] Relative risk is defined such that the probability of an event occurring to a given member of the treatment group is $r.\tau$, where τ is the probability of an event occurring to an otherwise identical member of the control group (baseline risk). If the treatment and control groups were both part of an otherwise perfectly homogenous population, an efficient and unbiased estimate of the relative risk could be obtained simply by calculating the ratio of the in-sample incidence rate in the treatment and control groups. Such homogeneous populations are, of course, vanishingly rare except in designed experiments. In almost all other cases, treatment and control groups differ systematically in terms of the relative frequency of occurrence of characteristics which themselves influence the probability of the event (e.g. income, occupation, place of residence).

In order to avoid calculating biased estimates of the relative risk, the researcher needs to ensure that treated observations are compared only to

about the relative propensity of female-owned firms exporting.

[16]When a conditional logit estimator is used, the odds ratio cannot be converted to a relative risk, nor can marginal effects be estimated, because the baseline risk for each strata is not estimated.

[17]The alternative approaches tested included generalized linear models with Poisson or Binomial distribution assumptions, an ordinary logit estimator, and propensity score matching. Code and Stata output available on request from the authors.

[18]For our purposes and most social science applications these terms are inter-changeable.

[19]Note that 'treatment' variable of interest is defined by the researcher. For example the treatment group could be city residents and control group rural residents.

like controls. This can be achieved through the definition of matched subpopulations, or strata, within the treatment and control groups.[20] Unbiased estimates of relative risk can then be calculated separately for each of the strata. The challenge, then, is how best to combine the information from each strata to produce a single estimate of relative risk for the population.[21]

Let us index the strata defined by the researcher $k = 1, ..K$. The key to our approach is to condition out the strata effects (which are incidental since we are only interested in estimating the effect of our treatment variable). To do this, we write the probability of an event occurring to a member of the treatment group in strata k, conditional on the event having occurred to some member of that strata. To begin, we assume that events in both treatment and control groups are independent Bernoulli draws.

Let the probability of an event occurring in a single draw from the control group in strata k be τ_k. Using the definition of relative risk, the corresponding probability in the treatment group is then $r\tau_k$. Taking a random draw from the population in strata k, the probability of it being a member of the control group is $\frac{n_{k_0}}{n_{k_1}+n_{k_0}}$ where n_{k_0} and n_{k_1} are the number of individuals in within strata k, for control (0) and treatment (1) groups respectively. Similarly, the probability that a member of the treatment group is drawn is $\frac{n_{k_1}}{n_{k_1}+n_{k_0}}$. Given the independence assumption, it thus follows that the probability of a member of the control population that experienced the event being drawn from the population of strata k is $\tau_k \frac{n_{k_0}}{n_{k_1}+n_{k_0}}$. Similarly, the probability of a member of the treatment group which experienced an event being drawn is

$$r\tau_k \frac{n_{k_1}}{n_{k_1} + n_{k_0}}. \tag{1}$$

The probability of drawing an event from either control or treatment populations is thus the sum

$$\frac{n_{k_0}}{n_{k_1} + n_{k_0}}\tau_k + \frac{n_{k_1}}{n_{k_1} + n_{k_0}} r\tau_k. \tag{2}$$

[20] In the case of survey data, strata will almost always already have been defined, but the researcher is not restricted to using those defined by the survey instrument. See Section 4.1 for an example of strata definition.

[21] We assume that the true relative risk, r, is constant across the strata. If the researcher does not believe that the relative risk is constant across the strata, they should not be trying to estimate a combined measure for the population under study. The assumption of constant relative risk can, of course, be tested.

Again using the independence of events in the treatment and control groups, the probability of an event having occurred in the treatment group of strata k, conditional on an event having occurred to some member of strata k is the ratio of the two marginal probabilities given by equations 1 and 2. This simplifies to

$$\frac{rn_{k_1}}{rn_{k_1} + n_{k_0}}. \tag{3}$$

Let c_{k_i} be the number of events within strata k in group $i = 0, 1$ and C_k be the total/combined number of events in both treatment and control (0 and 1) groups in strata k. Assuming the events are independent, c_{k_1} is the sum of successes from C_k repeated, independent, binary draws (Bernoulli trials), each with a probability of success given by equation 3. It is well-known that in this case c_{k_1} will have a Binomial distribution[22]

$$c_{k_1} \sim B\left(C_k, \frac{rn_{k_1}}{rn_{k_1} + n_{k_0}}\right). \tag{4}$$

Hence the conditional likelihood of the observed data is given by

$$\mathcal{L}(r; C_1 ... C_K) = \prod_{k=1}^{K} \binom{C_k}{c_{k_1}} p^{c_{k_1}} (1-p)^{C_k - c_{k_1}} \tag{5}$$

where $p \equiv \frac{rn_{k_1}}{rn_{k_1}+n_{k_0}}$ and

$$\binom{C_k}{c_{k_1}} = \frac{C_k!}{c_{k_1}!\,(C_k - c_{k_1})!}.$$

Simplifying the first order condition for maximization of the log of the conditional likelihood gives us our estimating equation

$$\frac{d}{dr}\ell(r) = r\left(\sum_{k=1}^{K} \frac{n_{k_0} C_k}{rn_{k_1} + n_{k_0}} - \sum_{k=1}^{K} (C_k - c_{k_1})\right) = 0. \tag{6}$$

Assuming that the true relative risk is not equal to zero, our estimator $\hat{r}$ is the solution to

$$\left(\sum_{k=1}^{K} \frac{n_{k_0} C_k}{rn_{k_1} + n_{k_0}} - \sum_{k=1}^{K} (C_k - c_{k_1})\right) = 0. \tag{7}$$

[22]See for example Rao (1952).

Solving equation 7 we obtain an iterative formula for the conditional likelihood estimator of relative risk (CLERR),

$$\hat{r} = \frac{\sum_{k=1}^{K} (c_{k_1} \frac{n_{k_0}}{n_{k_1}})/(\hat{r} + \frac{n_{k_0}}{n_{k_1}})}{\sum_{k=1}^{K} c_{k_0}/(\hat{r} + \frac{n_{k_0}}{n_{k_1}})}. \tag{8}$$

Although we do not have a closed form solution for $\hat{r}$, iterative estimation converges quickly to a unique solution because the estimating equation is analytic and monotonically increasing in $\hat{r}$, and its derivative is monotonically decreasing in $\hat{r}$.

3 Properties of the CLERR

In this section we discuss the favourable set of statistical properties of the CLERR, for both small and large sample assumptions. In small samples we show that the CLERR is unbiased and that our estimating equation is the optimal one (conditional or otherwise). The asymptotic properties of the CLERR are similar to those typical of maximum likelihood estimators, namely consistency, asymptotic normality, achievement of the Cramer-Rao lower bound and, hence, efficiency.

The impressive array of attributes of the CLERR can be traced back to the underlying Bernoulli distributions of the events in both treatment and control groups and in each strata. This underlying distribution, combined with the assumption of independence of events, means that the event counts in each group and strata are binomially distributed.[23] Since the Binomial family of distributions is complete, the combined sum of events in treatment and control groups within a strata is binomially distributed, as are the sums of events (within groups or combined for both groups) over all strata.

In order to keep the exposition as succinct and accessible as possible, we avoid highly formal presentation of the proofs below. In essence, the results follow from the fact that the baseline risk is bounded at one, ensuring that the resulting Binomial distributions will also be 'well-behaved' in the sense

[23] Note that the estimating equation would not change if we assumed that the underlying distributions of events in both treatment and control groups were Poisson, not Binomial. Thus the CLERR could be thought of as a "conditional Poisson" estimator, generalizing the "fixed effects" Poisson estimator of Hausman et al. (1984) to allow for variable and unequal numbers of individuals in the treatment and controls groups in each strata.

that they meet the required regularity conditions, the favorable asymptotic properties of conditional maximum likelihood estimators as demonstrated by Andersen (1970). Additionally, since the Binomial family is a member of the exponential family of distributions, we will be able to make use of another of Andersen's (1970) results to show asymptotic efficiency. Finally, it is easy to prove that the CLERR has an additional property which is less common among maximum likelihood estimators, namely, unbiasedness. We begin with this proof.

3.1 Unbiasedness

To show that the CLERR is unbiased, we first take expectations of both sides of the estimating equation (7), conditional on the vector of observed total numbers of events in each strata, C. The left hand side is

$$E_C\left[\left(\sum_{k=1}^{K}\frac{n_{k_0}C_k}{\hat{r}n_{k_1}+n_{k_0}}-\sum_{k=1}^{K}(C_k-c_{k_1})\right)\right]. \tag{9}$$

Noting that $\hat{r}$ and the c_{k_1} are the only random variables, and $E_C\left[c_{k_1}\right]=C_k\left(\frac{rn_{k_1}}{rn_{k_1}+n_{k_0}}\right)$, with trivial algebra 9 becomes

$$E_C\left[r-\hat{r}\right]K\sum_{k=1}^{K}C_k n_{k_1}n_{k_0}. \tag{10}$$

Since all other terms are strictly positive, and the expectation of the right hand side of 7 is zero, we have $E_C\left[r-\hat{r}\right]=0$. Since this is true for all C, by the law of iterated expectations the unconditional expectation is also zero, $E\left[r-\hat{r}\right]=E[E_C\left[r-\hat{r}\right]]=0$. The CLERR is unbiased.

3.2 Optimality of the Estimating Function

Godambe (1976, p. 277) shows that under certain conditions the maximum conditional likelihood equation "provides the optimum estimating equation, the criterion of optimality being independent of conditioning." In the current section we demonstrate that our case meets these conditions and that equation 6 is, therefore, the optimum estimating equation.

Godambe (1976) considers the case in which for $\theta=(\theta_1,\theta_2)\in\Omega$, the one dimensional $\theta_1\in\Omega_1$ is the parameter of interest, and θ_2 is an unknown nuisance parameter vector. In this notation, the optimum estimating equation, g^*, among the set of unbiased estimating functions, $g(x, \theta_1)$, is that which minimizes the expected value of the variance of the standardized estimating function, $E_\theta \left\{g/E_\theta(\partial g/\partial\theta_1)\right\}^2$.

Godambe's Theorem 3.2 states that the conditional maximum likelihood equation will be the unique optimum estimating equation when the following conditions 1 and 2 and set of regularity conditions hold.

Condition 1.[24] The conditional frequency function of x given t depends on θ only through θ_1, that is

$$p(x, \theta) = f_x(x, \theta_1)h(t, \theta) \tag{11}$$

where $h(t, \theta)$ is the frequency function of t, a statistic of x.

Condition 2.[25] If $P^t_\theta = P^t_{\theta_1,\theta_2}$ is the distribution of t defined by 11 and

$$\mathcal{P}^t_{\theta_1} = \left\{P^t_{\theta_1,\theta_2} : \theta_2 \in \Omega_2, \theta_1\right\}, \tag{12}$$

then the class $\mathcal{P}^t_{\theta_1}$ is complete for every fixed θ_1 in Ω_1.

Godambe's regularity conditions are that for the probability density p, defined with respect to a measure μ: (a) Ω_1 is a real interval; (b) $\partial p\ \partial\theta_1$ exists $(\theta \in \Omega)$; (c) $\int pd\mu$ is differentiable under the integral sign with respect to θ_1 $(\theta \in \Omega)$; and (d) conditions corresponding to (b) and (c) are also satisfied for the frequency functions f_t and h. For the estimating function, g, Godambe assumes $(for \theta \in \Omega)$ (i) for every fixed θ g is measurable with respect to μ; (ii) $E_\theta[g] = 0$ (i.e. the estimating equation is unbiased); (iii) $\partial g/\partial\theta_1$ exists ; (iv) $\int gpd\mu$ is differentiable under the integral sign with respect to θ_1; (v) $E_{theta}(\partial g/\partial\theta_1)^2 > 0$ (vi) $\int g(\partial log f_t/\partial\theta_1)pd\mu$ and $\int g(\partial log h/\partial\theta_1)pd\mu$ exist.

Once we translate the notation, it is easy to see that Condition 1 is met for the CLERR. In our case, the random variable x is the matrix of event counts (with elements c_{k_i}) for treatment and control groups $(i = (0, 1))$ in each strata $(k = 1, ...K)$. Similarly, in our case, Godambe's statistic t is the vector of the total number of events in both treatment and control groups in each strata (with elements C_k). Let τ be the vector of parameters describing the baseline risk in each strata. With the assumption of independent events, we know $p(c, r, \tau) = f(c, r, \tau|C)h(C, r, \tau)$. Furthermore, from equation 4 we

[24] Godambe Assumption 1.1.

[25] Godambe Assumption 3.4.

know that the distributions of the c_{k_i} are independent of τ.[26] Hence we can write $p(c, r, \tau) = f_c(c, r)h(C, r, \tau)$ and Godambe's first condition is met.

To show that Condition 2 is met, we must first identify the distribution of our summary statistic, C (i.e. determine the form of P^t_θ). To do this, we return first to the underlying Bernoulli trials whose sum within a strata is C_k. Recall that the probability of success for a random draw taken from strata k is given by equation 2. Since our statistic C_k is the sum of successes from these Bernoulli trials, it will have a Binomial distribution with count parameter equal to $N_k = n_{k_1} + n_{k_0}$ and probability of success given by 2. One of the well-known properties of Binomial distributions is that they form a complete family (Lehmann and Scheff, 1950).[27]

In the same manner, it is easy to show that the marginal distributions of c_{k_1}, the count of events in the treatment group within each strata, are also Binomial. Due to the completeness property of the family of Binomial distributions, the weighted sum of these counts over the set of strata will also be binomially distributed. Given this, we can rely on the well-studied properties of Binomial distributions and conditional likelihood estimators to assert that the frequency functions p, f_t and h and the estimating equation g satisfy the regularity conditions assumed by Godambe.

3.3 Consistency and Asymptotic Normality

Andersen (1970) shows that under a set of fairly weak assumptions, conditional likelihood estimators which make use of minimal sufficient statistics for the incidental parameters are consistent and asymptotically normal, even if the number of incidental parameters increases proportionally to the number of observations. Furthermore, in Section 9 of that paper, Andersen shows that exponential-family, pair-wise comparison estimators like the CLERR do indeed condition on a set of minimally sufficient statistics and meet all of the other required regularity conditions.

[26]Technically equation 4 gives only the conditional distribution of the event counts for the treatment group, but it is trivial to show that the distribution of event counts for the control group is also independent of τ.

[27]More generally the Binomial distribution with known count is part of the single-parameter exponential family of distributions and thus $h(C, r, \tau)$ possesses all the favorable properties of this family.

3.4 Asymptotic Efficiency

To show Asymptotic Efficiency we rely on Theorem 7 of Andersen (1970, p.296). This theorem applies specifically to conditional maximum likelihood estimators of parameters of distributions in the exponential family, like the CLERR. It says if the regularity conditions required for the estimator to be consistent and asymptotically normal are met: Then t_k is weakly ancillary with respect to θ_1 for all k if and only if σ_K^2 is equal to the lower bound for the asymptotic variance.

Weak ancilliarity is defined by Andersen as follows (p. 296). Consider the family Π_t of marginal distributions of the summary statistic, t. t is weakly ancillary with respect to θ_1 if for any given set of values $(\theta_1^0, \theta_2^0) \in (\Omega_1, \Omega_2)$ and for any other value $\theta_1 \in \Omega_1$ there exists a point $\theta_2 = \theta_2(\theta_1) \in \Omega_2$ such that the resulting marginal distributions of t are identical. The intuition behind the result is that the summary statistic t cannot give us any information about θ_1 if θ_1 is not already known. Thus conditioning on t does not amount to "throwing away" any information and the conditional likelihood estimator retains the efficiency properties of a full maximum likelihood estimator.

We have already established that the marginal distributions of the elements C_k are Binomial with count parameter equal to $N_k = n_{k_1} + n_{k_0}$ and probability of success given by 2. Denote the evaluation of equation 2 at (r^0, τ_2^0), P^0. It is trivial to show that we can always find a $\tau(r)$ such that equation 2 evaluated at $(r, \tau(r))$ equals P^0.

Hence C_k is weakly ancillary with respect to r for all k and the CLERR is asymptotically efficient.

3.5 Asymptotic Standard Error

The asymptotic variance of the CLERR equals the inverse of the negative of the Hessian of the conditional likelihood function evaluated at the maximum likelihood estimates. Thus the asymptotic standard error of the CLERR is

$$\sigma = \left\{ \frac{1}{r} \left(\sum_{k=1}^{K} \frac{C_k}{r + n_{k_0}/n_{k_1}} (1 - \frac{r}{r + n_{k_0}/n_{k_1}}) \right) \right\}^{-\frac{1}{2}} \tag{13}$$

which can be estimated using the observed values of the C_k and the estimated value of r. We can then construct Wald tests of the hypothesis $r = 1$ using the sample estimate from equation 13.

4 Applying the CLERR: Female Owners and Export Propensity

We now turn to demonstrating the application of the CLERR, and comparing its results to those obtained from the conditional logit estimator. Our empirical question is how the propensity of firms to engage in exporting is affected by whether they have at least one female owner. Specifically, we estimate the relative risk of exporting for firms with female owners (female-owner firms) compared to those with only male owners.

There is a large literature on the firm-level determinants of export participation (Katsikeas et al., 2000) and a similarly substantial literature on the firm-level impacts of females in leadership positions within firms (Carter et al., 2003; Erhardt et al., 2003; Nina Smith et al., 2006; Dezs and Ross, 2012). Our findings have relevance for both of these literatures. They additionally speak to the substantial literature on the gender-distribution of the benefits from globalization. We are aware of only one previous paper which looks specifically at the question of gender and export propensity (Orser et al., 2010). Orser et al. (2010) apply a (unconditional) logit model to a dataset of Canadian small- and medium-sized enterprises. Although the authors claim to find that female majority-owned firms were significantly less likely to export than firms owned by men, our reading of their results tables is otherwise. Furthermore, their choice of control variables and interaction terms for their model demonstrates exactly the sort of researcher subjectivity issues which using the CLERR avoids.

One of the most robust findings from the behavioural economics literature is that women are on average more risk adverse than men (Borghans et al., 2009; Eckel and Grossman, 2008). This relative risk aversion is also found to affect female firm manager's decisions (Khan and Vieito, 2013). Of particular relevance to our question, Downing (1991) have shown that female managers devote a relatively large proportion of available resources to diversification rather than expansion.

Most people's initial intuition would probably be to assume that exporting is something more likely to be avoided by risk-averse managers. As Hirsch and Lev (1971, p.270) note in their seminal article, "Folklore has it that foreign markets are more risky than domestic markets because of political, economic, and social instability abroad." These authors then go on to show both theoretically and empirically, that this common assumption is wrong. In short, they show that exporting can be understood as diversification strategy, whereby, even if the new market is riskier, the overall sales risk can be lowered. Furthermore, there is evidence that small and medium-sized enterprises in developing countries are particularly likely to use exporting as a means of diversification (Aw and Batra, 1998). According to Aw and Batra (1998)[p.313] "Our findings indicate that diversification need not be solely a large firm phenomenon as observed in developed countries. Among small and medium firms, the most common form of diversification consists of diversifying into a different geographical market." This finding is particularly relevant in our case, because the Enterprise Survey Data is collected exclusively in developing and transition countries.

Taken together, the existing literature suggests that female-owner firms in developing and transition countries will be more likely to export than their all-male counterparts.

4.1 Data and Strata Definition

The World Bank's Enterprise Survey (WBES) collects data on a wide variety of topics[28] from key manufacturing and service sectors in every region of the world since 2006.[29] We observe 105,275 firms from 135 different countries[30], 583 locations within countries[31] and 552 sectors, from the transitioning and developing world. Table 1 shows the proportions and number of observations in the WBES Sample grouped by our variable of interest: Female Ownership, that is if "any of the owners are female" (Question b4 in the survey), which was answered with "yes" by 32.7%.

Table 1 shows there are a number of significant differences between male-only owned firms and firms with female ownership. While the actual differences are mostly small, we find a higher proportion among female owned firms engage in direct exporting (18% vs. 16% among the male-only firms), are in the top size quintile, operate as Public Limited or Limited Liability

[28]The topics covered in Enterprise Surveys include infrastructure, trade, finance, regulations, taxes and business licensing, corruption, crime and informality, finance, innovation, labor, and perceptions about obstacles to doing business.

[29]See http://www.enterprisesurveys.org/data for more general details.

[30]We have 192 country-years in the sample. The WBES is aiming to eventually provide a panel of all countries, at the moment a separate short panel data set exist for only a subset of countries.

[31]These are selected regions within a country the WBES covers, which represent the largest centers of production and business enterprise.

Companies or in a Limited Partnership and have government ownership. On the other hand a higher proportion among male-only owned firms are in the 2nd or 3rd firm size quintile, operate as General Partnerships or Sole Proprietorship, are subsidiaries and have foreign Ownership in firm. This makes it clear that, in order to answer the question if female owned firms are more likly to export, we need to control for all these differences that exist between the two groups of ownership. We will do so using all these variables to stratify the sample when estimating the CLERR.

We use three alternative strata definitions in our analysis. The least conservative of these defines strata as the intersection of country, industry and size classification[32] and creates around 8,000 non-empty strata. The next strata adds to this a distinction by legal status of the firm, resulting in around 16,000 non-empty strata. Finally, our most conservative strata definition replaces country of location with the exact town or city of location, creating around 30,000 non-empty strata[33]. Thus our most conservative specification compares the propensity of firms with some female owner with that of male-only owned firms of the same legal status, in the same industry, same size category and same town or city of operation. In contrast to many treatment effects estimators, we make clear in our results presentation the consequences of changing strata definition for the number of identified strata, number of identified observations, and number of identified, treated observations.

4.2 Results

Table 2 summarises the results of our initial empirical application. Columns 1-3 of Table 2 represent increasingly conservative strata specifications. The top panel of Table 2 shows how the relative risk, standard error, and p-value for our CLERR estimates vary across the specifications. The middle panel of the table does the same for conditional logit estimates. The bottom panel

[32]The WBES uses a firm size variable with three categories (Number of employees: 5-19; 20-99; >100). In order to create a finer categorization, and to illustrate how continuous variables can be used for stratification, we define five size categories by dividing log number of full time employees into quintiles (Mean number of employees in each quintiles: 6; 11; 20; 51; 427).

[33]The theoretical total number of strata generated with these variables is $552 * 583 * 5 = 1,609,080$, which underlines how detailed this last stratification really is, but not in every location exists every type of industry, nor has every industry very small as well as very large firm, and all in between. The number of non-empty strata is therefor much smaller, but we will see in the results tables that a majority of all observations are still matched.

shows the total number of strata, identified strata, identified observations, and identified, treated observations for each stata definition.[34]

From an economic perspective, the main conclusion from Table 2 is that the relative propensity of firms with some female owner(s) to export is never less than that of all-male owned firms. The relative risk estimated by the CLERR ranges from 1.048 in column 1, to 1.011 in column 2, indicating that female ownership is associated with a relative risk increase for exporting of between 1.1 and 4.8%. Controlling for the exact location of the firm (column 3) actually raises the estimated relative risk slightly to 1.018, though the difference between the estimates in columns 2 and 3 is by no means statistically significant. Importantly also, the p-statistics in columns 2 and 3 show that the relative risk is not statistically significantly different to 1.

The middle panel of Table 2 shows that the direction and statistical significance of the effects estimated by the conditional logit estimator are similar to those estimated by the CLERR. The magnitude of the effects estimated by the logit model, however, differ in economically significant ways from the CLERR estimates. If we were to rely on the conditional logit estimates, we would overestimate the effect of female ownership on relative export propensity by around 80%.[35]

Turning now to the bottom panel of Table 2, we see that with the first stratification of column 1 we end up with 7,773 non-empty strata of which 4,964 include both female and male-only firms. Of the total 105,275 observations in the sample we can use over 80%, namely 84,481, identified firms in the estimation of which 29,057 have some female ownership. We observe that the increasing stringency of strata definition across columns 1-3 results in a reduction of roughly one third of the total number of identified observations, and around a quarter of identified, treated observations. This reporting makes transparent the trade-offs involved in using increasingly conservative strata specifications.

Table 2 reports results for the full sample of (identified) firms in the dataset. This sample includes firms which are subsidiaries of other firms, firms which have foreign ownership greater than 10% and firms with government ownership greater than 10%. We posit that owners of these firms

[34]Note, because both estimators condition out strata effects, these values are all the same for the CLERR and the conditional logit estimator.

[35]Based on the conditional logit model one could conclude that female ownership is associated with a relative risk increase of between 1.9 and 8.2%, compared to between 1.1 and 4.8% using the CLERR estimates.

have less control and bear less personal risk than owners of fully independent firms. For this reason, we report in Table 3 the results of the analysis of the sub-sample of firms which do not have government ownership, foreign ownership or any other parent firm.

The CLERR estimates in Table 3 show that independent, female-owner firms are significantly more likely to export than their all-male counterparts. The estimates suggest that all-female ownership increases a firm's relative risk of exporting by between nine and thirteen percent. As was the case for Table 2, the conditional logit results show the same pattern of statistical significance as the CLERR results, but here again, the assumption that the odds ratio can be used to approximate relative risk would lead to a substantial overestimate of the latter.

Comparison of the results in Tables 3 and 2 suggests that the relative propensity of female-owned firms to export is higher where owners are more likely to have more control and bear more risk. This same trend appears evident in Table 4. The columns in Table 4 present the results for subsamples of the data according to the legal organization of the firm. Moving from column 1 to 5 we approximately move in order of increasing control and financial risk for the owners. Columns 1-5 report results for the subsamples of: Public Limited Company (PLC); (privately held) Limited Liability Company (LLC); Limited Partnership (LP); General Partnerships (GP); Sole Proprietorship (SP). In all subsamples female-owner firms are more likely to export than male-only firms, however this difference is not statistically significant at the 10% level for publicly listed firms. The lowest estimated relative risk is for private limited liability firms, though due to a smaller standard error this estimate is still significantly different from one at the 5% level of significance. The relative propensities estimated for both types of partnership are also significantly different from one at least the 5% level of significance. These relative propensities are also unarguably economically significant. All-female ownership increases the relative propensity to export by around 21% for limited partnerships and around 29% for general partnerships. This latter figure is narrowly higher than the figure of 28% for sole proprietorships. Overall in Table 4, legal forms which are associated with increasing risk and responsibility for the owners tend to display a higher relative propensity of female-owned firms to export.

The result for sole proprietorships in Table 4 is also important for another reason. Up until this point, it had not been possible to distinguish solely-female owned firms. As a result, none of the earlier estimates could distinguish whether female owners are more likely to export than male owners, or whether mixed teams are more likely to export than single-gender teams. The results in column 5 suggest that it is, indeed, the presence of a female, rather than the mixed-gender team which is associated with increased export propensity.

Returning now to the question of estimator performance, we consider the results from the conditional logit estimator in the middle panel of Table 3. As has consistently been the case, using the conditional logit to estimate relative risk would lead to a substantial overestimate of the excess propensity of female-owner firms to engage in exporting. In Table 4, this overestimation is sometimes more than 100%. For example, the increase in relative propensity to export due to female ownership of public limited companies (PLCs) estimated by the CLERR is approximately 15%, while based on the conditional logit we would conclude this figure was around 34%.

5 Conclusion

The major contribution of this paper is methodological. We have introduced a novel estimator of relative risk of an event, and explained how it can be used to estimate treatment effects for binary dependent variables, and proven many of its superior statistical properties. Aside from being efficient and unbiased, our conditional likelihood estimator of relative risk (CLERR) is fast and easy to understand and apply. It is also flexible enough to be used either on individual-level (micro) data, or on event counts aggregated at the strata level. This latter property is an important advantage where privacy or cost concerns preclude access to unit-level data.

We see several avenues along which our work could readily be extended. The most obvious of these is to allow the relative risk to be a linear exponential function of explanatory variables — analogous to the full conditional logit estimator (McFadden, 1974) or the Poisson fixed effects estimator (Hausman et al., 1984). This extension would allow researchers the flexibility to control for some confounding variables parametrically. Such flexibility is useful when valuable information is lost in summarizing continuous variables as categories for stratification, or when there are too few observations relative to the dimensionality of the necessary control variables.

We demonstrated the application of the CLERR using the World Bank's Business Environment Survey data. The application illustrated the nontrivial nature of the biases associated with using the conditional logit to estimate relative risk, even for large sample sizes. Future work using Monte Carlo methods will enable us to distinguish the extent to which the observed differences between the conditional logit and CLERR results are due to bias in the conditional logit estimates or genuine differences between the true odds ratio and the relative risk.

Our analysis of the World Business Environment Survey data provided an interesting finding in its own right; namely, that firms with female owners are more likely to export than all-male owned firms. This finding is consistent with the idea that female owners utilize exporting as a means of risk-reducing diversification. It suggests both that exporting can help female entrepreneurs, and that female entrepreneurs can help expand countries' exports. We leave it to future research to explore formally the causal mechanism which drives the relationship between female ownership and exporting.

6 Appendix: Legal Status Definitions in World Bank Survey

According to the Questionnaire Notes[36] from the World Bank's Enterprise Survey: A firm's legal status is first determined by whether it has publicly or privately held shares. Partnerships or sole proprietorships implicitly have privately-held shares. After a determination is made as to whether shares are held publicly or privately, a firm's legal status is defined by the extent of the liability. Sole proprietorships and simple partnerships are the only entities with unlimited liability. All firms should fit into one (and only one) of these categories.

[PLC] If a firm's shares are publicly traded, it is a publicly listed company. A publicly listed company is also a limited liability company.

[LLC] A privately held, limited liability company is a firm that is owned by partners or shareholders for whom their claims over the firm are not publicly traded. They may or may not be traded privately.

[LP] Limited partnership is a type of business that includes one or several general partners and one or more limited partners who invest capital into the partnership, but do not take part in the daily operation or management of the business. The limited partners limit their amount of liability

[36] Available at www.enterprisesurveys.org

to the amount of capital invested in the partnership. The general partners personally shoulder all debts and obligations of the partnership. Business operations are governed, unless otherwise specified in a written agreement, by majority vote of voting partners. Limited liability partnerships are separate legal entities that provide liability protection for all general partners as well as management rights in the business.

[GP] A partnership allows two or more people to share profits and liabilities, with or without privately held shares. In a partnership, the parties could be individuals, corporations, trusts, other partnerships, or a combination of all of the above. The essential characteristic of the partnership is the unlimited liability of every partner.

[SP] A sole proprietorship is a business owned and operated by one individual, physical or juridical person. A juridical person can be aggregates of persons.

[Other - not used]Cooperatives should be designated as Other. The form of legal status must be specified in writing by the enumerator on the survey instrument.

Table 1: Summary Statistics of the World Bank's Enterprise Survey by Treatment Variable

Variable Name	Definition (Question No. in Survey)	By Ownership Male-Only	Female	
		Proportion (No. Observations)		z-score
Exporter	Equals 1 if Exports are positive. (Q d3c)	0.16 (10,370)	0.18 (5,708)	-8.13
Very Small	Equals 1 if in 1st quintile of log of total number of full time employees, adjusted for temporary workers. (Q l1)	0.19 (11,975)	0.18 (5,698)	1.37
Small	Equals 1 if in 2nd quintile, see above (Q l1)	0.20 (12,830)	0.19 (5,938)	3.39
Medium	Equals 1 if in 3rd quintile, see above (Q l1)	0.20 (13,101)	0.19 (6,057)	3.51
Large	Equals 1 if in 4th quintile, see above (Q l1)	0.22 (14,081)	0.22 (6,879)	-0.50
Very Large	Equals 1 if in 5th quintile, see above (Q l1)	0.19 (12,475)	0.21 (6,718)	-7.68
PLC	Equals 1 if form of ownership is Public Limited Company (Q b1)	0.04 (2,458)	0.06 (1,863)	-14.98
LLC	Equals 1 if (privately held) Limited Liability Company (Q b1)	0.44 (28,413)	0.53 (16,607)	-26.15
LP	Equals 1 if Limited Partnership (Q b1)	0.07 (4,235)	0.09 (2,688)	-11.34
GP	Equals 1 if General Partnerships (Q b1)	0.08 (5,101)	0.07 (2,321)	2.67
SP	Equals 1 if Sole Proprietorship (Q b1)	0.36 (23,496)	0.23 (7,090)	42.84
Subsidiary	Equals 1 if Establishment is part of a larger firm (Q a7)	0.16 (9,823)	0.14 (4,221)	8.86
FDI	Equals 1 if percent of foreign Ownership in firm is larger than 10%. (Q b2b)	0.09 (5,896)	0.08 (2,543)	5.09
GOV	Equals 1 if percent of Government Ownership is larger than 10%. (Q b2c)	0.01 (611)	0.02 (588)	-12.21

Notes: Column 5 are the z-score results of testing for each variable if the proportions are equal between Male-Only and Female Ownership. The total number of observation in our data set is 105,275.

Table 2: Female Ownership and Propensity to Export – CLERR and Conditional Logit Results

Treatment Variable:	Female Ownership		
Stratification by:	Country, Industry, Firm size	Country, Industry, Firm size, Legal status	City, Industry, Firm size, Legal status
	CLERR estimates		
Relative Risk	1.0476	1.0107	1.0179
Standard Error	0.0150	0.0156	0.0171
P-value	0.001	0.493	0.297
	Conditional Logit Estimates		
Relative Risk	1.0823	1.0185	1.0322
Standard Error	0.0261	0.0263	0.0295
P-value	0.001	0.478	0.268
	Strata Statistics		
Total number of strata	7,773	15,661	29,283
Identified Strata	4,964	6,806	8,272
Identified Observations	84,481	74,550	57,148
Identified, Treated Observations	29,057	26,373	21,921

Table 3: Female Ownership and Propensity to Export (Excluding Subsidiaries, Government or Foreign Owned Firms)

Treatment Variable:	Female Ownership		
Stratification by:	country, industry, size	country, industry, size, legal status	city, industry, size, legal status
	CLERR estimates		
Relative Risk	1.1311	1.0891	1.100
Standard Error	0.0204	0.0214	0.0237
P-value	0.000	0.000	0.000
	Conditional Logit Estimates		
Relative Risk	1.2111	1.1431	1.1678
Standard Error	0.0359	0.0364	0.0415
P-value	0.000	0.000	0.000
	Strata Statistics		
Total number of strata	7,022	13,129	24,177
Identified Strata	4,347	5,539	6,555
Identified Observations	63,297	55,548	41,518
Identified, Treated Observations	22,385	20,021	16,401

Table 4: Female Ownership and Propensity to Export by Legal Status of Firm (Excluding Subsidiaries, Government or Foreign-Owned Firms)

Treatment Variable:	Female Ownership				
Stratification by:	city, industry, size	city, industry, size	city, industry, size	city, industry, size	city, industry, size
Subsamples of:	PLC	LLC	LP	GP	SP
CLERR estimates					
Relative Risk	1.1517	1.0605	1.2148	1.2858	1.2790
Standard Error	0.1504	0.0263	0.1032	0.1370	0.0822
P-value	0.313	0.021	0.038	0.037	0.001
Conditional Logit Estimates					
Relative Risk	1.3398	1.1033	1.3372	1.4648	1.3947
Standard Error	0.3246	0.0453	0.1826	0.2620	0.1422
P-value	0.227	0.017	0.033	0.033	0.001
Strata Statistics					
Total number of strata	1,663	9,781	2,328	2,573	7,779
Identified Strata	209	3,523	488	461	1,874
Identified Observations	688	23,998	2,027	2,003	12,802
Identified, Treated Observations	327	9,905	934	838	4,397

Notes: As in the estimation in Table 2 Subsidiaries, Government or Foreign Owned Firms are excluded. The legal forms of ownership abbreviations above are as follows: Public Limited Company (PLC); (privately held) Limited Liability Company (LLC); Limited Partnership (LP); General Partnerships (GP); Sole Proprietorship (SP). Further details to the legal status definitions used in the survey can be found in the Appendix 2.

References

Andersen, E. B., Jan. 1970. Asymptotic Properties of Conditional Maximum-Likelihood Estimators. Journal of the Royal Statistical Society. Series B (Methodological) 32 (2), 283–301.
URL http://www.jstor.org/stable/2984535

Aw, B.-Y., Batra, G., May 1998. Firm size and the pattern of diversification. International Journal of Industrial Organization 16 (3), 313–331.
URL http://www.sciencedirect.com/science/article/pii/S0167718796010570

Barros, A. J., Hirakata, V. N., Oct. 2003. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Medical Research Methodology 3, 21.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC521200/

Borghans, L., Heckman, J. J., Golsteyn, B. H. H., Meijers, H., Apr. 2009. Gender Differences in Risk Aversion and Ambiguity Aversion. Journal of the European Economic Association 7 (2-3), 649–658.
URL http://onlinelibrary.wiley.com/doi/10.1162/JEEA.2009.7.2-3.649/abstract

Breslow, N., Jan. 1981. Odds ratio estimators when the data are sparse. Biometrika 68 (1), 73–84.
URL http://biomet.oxfordjournals.org/content/68/1/73

Breslow, N. E., Jan. 1984. Elementary Methods of Cohort Analysis*. International Journal of Epidemiology 13 (1), 112–115.
URL http://ije.oxfordjournals.org/content/13/1/112

Breslow, N. E., Day, N. E., 1980. Statistical methods in cancer research. International Agency for Research on Cancer, Lyon.

Carter, D. A., Simkins, B. J., Simpson, W. G., Feb. 2003. Corporate Governance, Board Diversity, and Firm Value. Financial Review 38 (1), 33–53.
URL http://onlinelibrary.wiley.com/doi/10.1111/1540-6288.00034/abstract

Coutinho, L. M. S., Scazufca, M., Menezes, P. R., Dec. 2008. Methods for estimating prevalence ratios in cross-sectional studies. Revista De Sade Pblica 42 (6), 992–998.

Cummings, P., 2009. Methods for estimating adjusted risk ratios. Stata Journal 9 (2), 175.
URL http://ageconsearch.umn.edu/bitstream/127328/2/sjart_st0162.pdf

Dezs, C. L., Ross, D. G., Sep. 2012. Does female representation in top management improve firm performance? A panel data investigation. Strategic Management Journal 33 (9), 1072–1089.
URL http://onlinelibrary.wiley.com/doi/10.1002/smj.1955/abstract

Diaz-Quijano, F. A., Feb. 2012. A simple method for estimating relative risk using logistic regression. BMC Medical Research Methodology 12, 14.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3305608/

Downing, J., Mar. 1991. Gender and the growth of microenterprises. Small Enterprise Development 2 (1), 4–12.
URL http://www.developmentbookshelf.com/doi/abs/10.3362/0957-1329.1991.002

Eckel, C. C., Grossman, P. J., 2008. Chapter 113 Men, Women and Risk Aversion: Experimental Evidence. In: Smith, C. R. P. a. V. L. (Ed.), Handbook of Experimental Economics Results. Vol. 1. Elsevier, pp. 1061–1073.
URL http://www.sciencedirect.com/science/article/pii/S1574072207001138

Erhardt, N. L., Werbel, J. D., Shrader, C. B., Apr. 2003. Board of Director Diversity and Firm Financial Performance. Corporate Governance: An International Review 11 (2), 102–111.
URL http://onlinelibrary.wiley.com/doi/10.1111/1467-8683.00011/abstract

Gart, J. J., Jan. 1978. The analysis of ratios and cross-product ratios of poisson variates with application to incidence rates. Communications in Statistics - Theory and Methods 7 (10), 917–937.

URL http://www.tandfonline.com/doi/abs/10.1080/03610927808827683

Godambe, V. P., Jan. 1976. Conditional likelihood and unconditional optimum estimating equations. Biometrika 63 (2), 277–284.
URL http://biomet.oxfordjournals.org/content/63/2/277

Greenland, S., Jan. 2000. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. Biostatistics 1 (1), 113–122.
URL http://biostatistics.oxfordjournals.org/content/1/1/113

Greenland, S., Robins, J. M., Mar. 1985. Estimation of a Common Effect Parameter from Sparse Follow-Up Data. Biometrics 41 (1), 55.
URL http://www.jstor.org/stable/2530643?origin=crossref

Greenland, S., Schwartzbaum, J. A., Finkle, W. D., Mar. 2000. Problems due to Small Samples and Sparse Data in Conditional Logistic Regression Analysis. American Journal of Epidemiology 151 (5), 531–539.
URL http://aje.oxfordjournals.org/cgi/doi/10.1093/oxfordjournals.aje.a010240

Hausman, J., Hall, B. H., Griliches, Z., Jul. 1984. Econometric Models for Count Data with an Application to the Patents-R & D Relationship. Econometrica 52 (4), 909–938.
URL http://www.jstor.org/stable/1911191

Hirsch, S., Lev, B., Aug. 1971. Sales Stabilization Through Export Diversification. The Review of Economics and Statistics 53 (3), 270.
URL http://www.jstor.org/stable/1937971?origin=crossref

Ho, D. E., Imai, K., King, G., Stuart, E. A., Jun. 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis 15 (3), 199–236.
URL http://pan.oxfordjournals.org/content/15/3/199

Katsikeas, C. S., Leonidou, L. C., Morgan, N. A., Jan. 2000. Firm-Level Export Performance Assessment: Review, Evaluation, and Development. Journal of the Academy of Marketing Science 28 (4), 493–511.
URL http://jam.sagepub.com/content/28/4/493

Khan, W. A., Vieito, J. P., May 2013. Ceo gender and firm performance. Journal of Economics and Business 67, 55–66.
URL http://www.sciencedirect.com/science/article/pii/S014861951300012X

Kupper, L. L., 2005. Matched Analysis. In: Encyclopedia of Biostatistics. John Wiley & Sons, Ltd.
URL http://onlinelibrary.wiley.com.virtual.anu.edu.au/doi/10.1002/0470011815.b2a03080/abstract

Lee, J., Tan, C. S., Chia, K. S., Aug. 2009. A practical guide for multivariate analysis of dichotomous outcomes. Annals of the Academy of Medicine, Singapore 38 (8), 714–719.

Lehmann, E. L., Scheff, H., 1950. Completeness, Similar Regions, and Unbiased Estimation: Part I. Sankhy: The Indian Journal of Statistics (1933-1960) 10 (4), 305–340.
URL http://www.jstor.org/stable/25048038

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Fronteirs in Econometrics. pp. 105–142.
URL https://elsa.berkeley.edu/reprints/mcfadden/zarembka.pdf

McNutt, L.-A., Wu, C., Xue, X., Hafner, J. P., May 2003. Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. American Journal of Epidemiology 157 (10), 940–943.
URL http://aje.oxfordjournals.org.virtual.anu.edu.au/content/157/10/940

Miller, J. C., Pras, B., Jan. 1980. The Effects of Multinational and Export Diversification on the Profit Stability of U. S. Corporations. Southern Economic Journal 46 (3), 792.
URL http://www.jstor.org/stable/1057148?origin=crossref

Nina Smith, Valdemar Smith, Mette Verner, Oct. 2006. Do women in top management affect firm performance?A panel study of 2,500 Danish firms. International Journal of Productivity and Performance Management 55 (7), 569–593.
URL http://www.emeraldinsight.com/doi/full/10.1108/17410400610702160

Orser, B., Spence, M., Riding, A., Carrington, C. A., Sep. 2010. Gender and Export Propensity. Entrepreneurship Theory and Practice 34 (5), 933–957.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6520.2009.00347.x/abstract