Essays on Partially Identifying Models and Their Applications

Dongwoo Kim

March 27, 2019

I, Dongwoo Kim, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Signature:
date:

Acknowledgements

Abstract

This thesis studies partial identification in discrete outcome models and their empirical applications. Chapter 1 investigates popular count data instrumental variable (IV) models. Many methods in the literature ignore the discreteness of count outcomes and thereby suffering from undesirable misspecification problems. To address this problem, a partially identifying count data IV model is developed. The model requires neither strong separability of unobserved heterogeneity nor a triangular system. Identified sets of structural features are derived. The size of the identified set can be very small when the support of an outcome is rich or instruments are strong. The proposed approach is applied to study effects of supplemental insurance on healthcare utilisation.

In Chapter 2, partial identification in competing risks models for discretely measured or interval censored durations are studied. These models are partially identifying because of 1) the unknown dependence structure between latent durations, and 2) the discrete nature of the outcome. I develop a highly tractable bounds approach for underlying distributions of latent durations by exploiting the discreteness and I investigate identifying power of restrictions on the dependence structure with no assumptions on covariate effects. Bounds are obtained from a system of nonlinear conditional moment (in)equalities. I devise a solution method that requires much less computational burden than existing methods. Asymptotic properties of bound estimators and a simple bootstrap procedure are provided.

Chapter 3 applies the proposed bounds approach in Chapter 2 to re-evaluate trends in cancer mortality by extending the "war on cancer" data studied in Honoré and Lleras-Muney (2006). I find substantial reduction in cancer mortality. Estimated patterns differ from the original findings. In another application, I investigate the effects of extended unemployment benefits on unemployment spells using data from Farber *et al.* (2015). Bound estimates support the original finding that extended benefits did not discourage active job seekers during and after the Great Recession.

Impact Statement

This thesis considers empirically relevant partially identifying models. The main feature of these models is that they require a weaker set of restrictions than point identifying models. Therefore, they are more robust to possible misspecification. Econometric analysis often relies on identifying restrictions which are not based on economic theory. Partially identifying models allow for applied researchers to relax such restrictions.

Count data and competing risks situations are widely studied in many applied fields. This thesis introduces easily implementable and computationally attractive partial identification approaches to those problems. Theoretical innovations in this thesis ease potential misspecification problems in the existing count data methods. When it comes to competing risks models, my new approach substantially mitigates computational difficulties from which existing methods in the literature have suffered and thereby make the models widely and easily applicable in applied studies with a valid inference procedure.

Outside academia, the proposed methods can be employed for programme evaluation as count and competing risks data are very common in many fields. Empirical applications in Chapter 1 and Chapter 3 illustrate the usefulness of the proposed approaches. Especially, newly revealed patterns in cancer mortality trends in the United States and counterfactual analysis on effects of further reductions in cancer and cardiovascular disease on overall life span provide useful insights on health policy. Another application to effects of unemployment insurance on unemployment spells supports the controversial argument that unemployment benefits do not distort economic efficiency. Dissemination of this research through journal publications is expected to boost the use of proposed methods.

Contents

List of Figures

List of Tables

Chapter 1

Partial Identification in Nonseparable Count Data IV Models

1.1 Introduction

This chapter introduces a new approach to count data instrumental variable (IV) models where explanatory variables are potentially endogenous and unobserved heterogeneity is nonseparable. The proposed approach is widely applicable in applied studies as many outcomes of interest are count-measured (Cameron and Trivedi (2013), CT13 henceforth).[1] Endogeneity is a common concern in economics. For instance, in the context of doctor visits, some observable characteristics can be correlated with unobserved factors. Suppose individuals self-reported their current health statuses. If they did not report whether they had private health insurance, explanatory variables such as occupations would be endogenous since having private insurance is probably correlated with health status as well as occupation.[2] If this is the case, the OLS estimator fails to deliver correct information about causal effects of interest. IV models are a usual ploy to cope with this problem.

I study identifying power of a single equation IV model for ordered outcomes introduced by Chesher (2010) and Chesher and Smolinski (2012) in the context of count data. Those two papers explore partial identification of the

---

[1] For instance, in health economics, the numbers of doctor visits and other types of health care utilization, occupational injuries and illnesses are all count outcomes. Other examples are widely found in labour and empirical IO, and even finance literature such as absenteeism in the workplaces, recreational or shopping trips, entry and exits from industries, mortgage prepayments and loan defaults, bank failures, patent registration in connection with industrial R&D, and frequency of airline accidents.

[2] People with risky occupations may want to have private insurance. Some workplaces for high skilled workers often provide private insurance to their employees.

structural function and derive the sharp identified set when either the outcome or the scalar endogenous explanatory variable is binary. In this chapter, the sharp identified set of the structural function is defined using the generalized IV model framework introduced in Chesher and Rosen (2017). The outcome and endogenous variables can be both non-binary so that sharpness is guaranteed regardless of supports of those variables.

Using this model for count outcomes is beneficial in the sense that widely used count data IV methods in the literature suffer from undesirable limitations as the discreteness of count data is ignored. I demonstrate that those approaches may deliver misleading information about the causal effects of interest. The proposed model explicitly accommodates the discreteness and hence is more robust to misspecification. In simulation studies, it is shown that identified sets delivered by the model always contain the true values of structural features, meanwhile other alternatives are in general inconsistent.

In empirical studies using partially identifying models, obtaining "tight" bounds is a primary concern. Therefore, it is important to learn what features of the distribution of data could deliver tight bounds (see Ho and Rosen (2015), Section 7.2 for detailed discussions). Count outcomes often have a richer support than other ordered outcomes depending upon the duration in which counts are aggregated. The richer support of the outcome in general leads to tighter bounds. I show that identified sets of structural features can be very small when the IV is strong or the support of the outcome is rich. A simple algorithm is introduced to compute identified sets.

Recent developments in the partial identification literature provide straightforward inference methods. Chernozhukov *et al.* (2013) develops a novel inference method on identified sets characterized by intersection bounds. Inference techniques on projections of high dimensional identified sets are introduced by Kaido *et al.* (2017) and Bugni *et al.* (2017). I employ the intersection bounds method to compute confidence regions for identified sets, thereby documenting a unified framework from identification to inference. An empirical application to effects of supplemental insurance on the number of doctor visits shows the usefulness of the proposed approach.

Two branches of count data IV methods are commonly used. The first is the control function approach.[3] This approach is widely used in applied studies but requires strong assumptions. Under some circumstances, the use of this approach is not recommended. For instance, the triangular structure rules out

---

[3]Terza *et al.* (2008) implements it in the context of count data models viz. 2 stage residual inclusion estimation (2SRI).

full simultaneity. Moreover, endogenous variables are required to be continuously distributed and so are the instruments unless the first stage regression is linear. The other branch is moment based approaches suggested in Windmeijer and Santos Silva (1997) (WS1997 henceforth) and Mullahy (1997). These approaches do not require a triangular system and parameters of interest are point identified. However, they ignore the discreteness of count data and separable errors absorb the discreteness. Consequently, conditional supports of separable errors depend on given values of explanatory variables. I show that no relevant instrument satisfies the independence condition if endogenous variables are discrete. Furthermore, point identified parameters explain not much about the DGP. Both approaches also restrict unobserved heterogeneity to be scalar. I consider multi-dimensional unobserved heterogeneity in Appendix A.

Relaxation of strong separability in incomplete models gives rise to partial identification. The importance of model specifications cannot be emphasized enough in applied economic studies. Applied researchers often impose simplifying assumptions which are not based on economic theories in order to make identification and estimation more tractable. In many cases, they become the primary source of misspecification. Partial identification approach tends to impose a minimal set of restrictions to extract useful information from data and hence it is less vulnerable to attacks on econometric assumptions. To my best knowledge, there has been no paper exploring partial identification in count data models.

The rest of this chapter is structured as follows. Section 2 shows potential problems of prevailing approaches in the literature. Section 3 introduces incomplete count data IV models with the nonseparable error and the characterization of identified sets. Section 4 demonstrates identified sets in numerical examples. Section 5 shows estimation and inference results on an empirical example. Section 6 concludes. All proofs are provided in Appendix A.

1.1.1 Notation

The notation in this chapter follows the convention in the literature. Upper and lower case letters A and a denote a random vector and its particular realization respectively. $\mathcal{R}_A$ denotes the support of A; $F_{A|B}(\cdot|b)$ denotes the conditional distribution function of A given the realization of a random variable B; The calligraphic font ($\mathcal{A}$) is reserved for sets and the sans serif font (A) is reserved for collections of sets. Y, X and Z denote a scalar count outcome, a vector of explanatory variables, and a vector of instrumental variables respectively. U denotes scalar unobserved heterogeneity.

1.2 Prevailing count data approaches

1.2.1 Control function approach

Terza *et al.* (2008) introduces the control function approach in the context of count data models. The model is specified as

$$Y \sim Poisson[\lambda(X,U)], \quad \lambda(X,U) = \exp(X'\beta + U) \tag{1.1}$$

$$X = g(Z) + V \tag{1.2}$$

$$U = \alpha V + e \tag{1.3}$$

where $P[Y \leq y; \lambda] = \exp(-\lambda) \sum_{m=0}^{y} \frac{\exp(\lambda^m)}{m!}$. Z is assumed to be independent of (e, V) and e and V are mutually independent. $E[\exp(e)]$ is normalised to 1. Then

$$E[\lambda(X,U)|X,Z] = E[\exp(e)|X,V]\exp(X'\beta + \alpha V) = \exp(X'\beta + \alpha V). \tag{1.4}$$

The last equality of (1.4) holds because e is independent of (Z,V) and X. V is identified by the second equation. The Poisson distribution for Y can be replaced by any other parametric count distribution.

This method is very tractable and widely used to deal with endogeneity but is somewhat restrictive in the sense that the recursive structure rules out full simultaneity (Koenker (2005), Section 8.8.3).[4] Moreover, the auxiliary first stage can be an additional source of misspecification. If the true function g is misspecified, then estimation results may be biased. Furthermore, X is generally required to be continuously distributed. Otherwise, the error term in the first stage is not separably identified.[5] The instrument Z is also required to be continuous unless the first stage is linear. Chesher (2005) shows set identification is possible when X is discrete and the error term is nonseparable. His method is not applicable if X is binary.

1.2.2 Moment based approaches

Moment based approaches are not reliant on the recursive structure. Suppose unobserved heterogeneity U is additively separable. Then the model is

---

[4]In simultaneous equation models, endogenous variables might affect each other. Therefore, the variation of Y can possibly induce the change of X. The recursive system rules out this relationship as Y is restricted to have no effect on X.

[5]If X is an ordered choice, then standard parametric models do not provide a single valued e given Z and X.

specified as follows.

$$Y = \exp(X'\beta) + U, \quad E[U|Z] = 0 \tag{1.5}$$

WS1997 shows that β is point identified by the moment condition

$$E[Z(Y - \exp(X'\beta))] = 0. \tag{1.6}$$

The generalized method of moments (GMM) estimator consistently estimates β. However, Mullahy (1997) points out that this specification treats X and U asymmetrically. Suppose now unobserved heterogeneity W is omitted characteristics. U is a regression error such that $E[U|X, W, Z] = 0$. Then the structural equation is written as

$$Y = \exp(X'\beta + W'\delta) + U = \exp(X'\beta)V + U, \quad \text{where} \quad V = \exp(W'\delta). \tag{1.7}$$

V is multiplicatively separable and X and V are treated symmetrically. Given $E[V|Z] = 1$, the moment condition

$$E\left[Z\left(\frac{Y}{\exp(X'\beta)} - 1\right)\right] = 0 \tag{1.8}$$

point identifies β as shown in Mullahy (1997). Two specifications (1.5) and (1.7) are observationally equivalent (see Wooldridge (1992)). These approaches only respect non-negativity of count outcomes and therefore they can be employed to investigate other non-negative outcomes such as birth weight.

The moment based approaches involve a fundamental problem when unobserved heterogeneity is interpreted as of economic interest. In econometric models with endogeneity, unobserved heterogeneity generally has a clear economic meaning. When it comes to returns to schooling, years of education (X) is supposed to be correlated with unobservable ability (U) which affects X as well as income (Y) for an individual. Therefore, a valid instrument Z is necessary to separately identify the causal effect of education on earnings from that of unobserved ability. Persuasive explanation about the relationship between Z and U should be presented as it is untestable.

Now suppose that a model specification per se restricts the distribution of U with which endowing U with economic interpretation is hard. If one cannot devise an economic example of such unobserved heterogeneity, then it would be also impossible to argue that there exists some good instrument Z. In the moment based models, the conditional support of U given $X = x$ is discrete

and varies with x and hence X and U cannot be independent by construction. This arises due to the attempt to fit the discrete outcome by a continuous function. U absorbs the discreteness of Y. However it is seldom justified to impose such discreteness on U. Can unobserved heterogeneity, whose discrete conditional support varies with X, be found in any economic example? How can one endow it with an economic meaning? These questions are hard to answer, even though these model specifications are very common in applied studies.

The more fundamental problem is that there exists no instrument which is independent of U but correlated with X if X is discrete and bounded. Suppose the model is $Y = \exp(\alpha + \beta X) + U$. The following proposition shows that existence of a good instrument is not guaranteed under the model (1.5).

Proposition 1. *Suppose that Y is a count outcome and X is a discrete and finite scalar explanatory variable i.e. $\mathcal{R}_X \equiv \{x_1, x_2, \cdots, x_n\}$. Under the model such that $Y = \exp(\alpha + \beta X) + U$, only a particular set of pairs (α, β), whose Lebesgue measure is zero, allows for the instrument Z being independent of U, but correlated with X.*

The true parameters are never known and the set of combinations of (α, β) under which X and Z are correlated is a measure zero subset of the parameter space. Therefore, the existence of a proper instrument is never assured. Even if the true parameters indeed lie on the particular set in Proposition 1, limited variation between certain values of X is allowed. The result in Proposition 1 is extended to the model (1.7). The additive error U is omitted here as it is redundant.

Proposition 2. *Suppose that Y is a count outcome and X is a discrete and finite scalar explanatory variable. Under the model $Y = \exp(\alpha + \beta X)V$, only a particular set of pairs (α, β), whose Lebesgue measure is zero, allows for the instrument Z being independent of U, but correlated with X.*

Remark 1. Conditional mean independence of U given Z is required to identify α and β. It is weaker than the strong independence condition in Proposition 1 and 2. However, in many applied studies, it is rarely justifiable to argue that Z satisfies conditional mean independence but is not independent of U. Most applied researchers argue that their instruments are completely independent of unobserved heterogeneity. (For example, see Angrist and Krueger (1991) and Angrist and Evans (1998).) However, this argument is impossible in the context of moment based models.

The moment based approaches ignore the discreteness of count outcomes. Even though parameters in the models are point identified, those do not necessarily tell about the underlying DGP. Needless to say, the moment based approaches do not work when unobserved heterogeneity is nonseparable.

1.3 Models with nonseparable error

An incomplete count data IV model is built using a threshold crossing specification introduced in Chesher and Smolinski (2012). Define $\mathcal{R}_Y$ as a subset of all non-negative integers i.e. $\mathcal{R}_Y \equiv \{0, 1, 2, \cdots\}$ and $y \in \mathcal{R}_Y$. $\mathcal{R}_Y$ is possibly unbounded. The model is

$$Y = h(X, U) = y \quad \text{if } p_y(X) \leq U < p_{y+1}(X) \tag{1.9}$$

where $p_0(X)$ is normalized to 0. U is normalized to be uniformly distributed on $[0, 1]$. The threshold functions $\{p_y(X)\}_{y=1}^{\infty}$ are objects of identification.

Suppose X is independent of U. Then it is reasonable to define the conditional distribution of Y given X as $p_{y+1}(X) = F_{Y|X}(y|X)$. Since $U \sim Unif(0, 1)$, the thresholds, $\{p_y(x)\}_{y=1}^{\infty}$, are all point identified by the conditional cdf of Y given $X = x$. Therefore, the full conditional distribution of Y given X is nonparametrically identified and it provides useful insight about the causal relationship between X and Y.

If X and U are not independent, thresholds functions are not point identified since the conditional distribution of U given X is not uniform. Suppose that X is binary i.e. $\mathcal{R}_X \equiv \{0, 1\}$ and $F_{U|X}(\cdot|X = 1)$ first order stochastically dominates $F_{U|X}(\cdot|X = 0)$. Then $F_{Y|X}(y|X = 1) \leq p_{y+1}(1)$ and $p_{y+1}(0) \leq F_{Y|X}(y|X = 0)$. Therefore without additional information, $\{p_y(0), p_y(1)\}_{y=1}^{\infty}$ are not identified. What one can identify are lower bounds for $\{p_y(1)\}_{y=1}^{\infty}$ and upper bounds for $\{p_y(0)\}_{y=1}^{\infty}$. Without the stochastic dominance assumption, one may be able to identify no-assumption bounds as in Manski and Pepper (2000).

The main question of this chapter is how to identify the threshold functions under the existence of instruments Z. As Chesher (2010) shows, point identification is generally not achievable in incomplete IV models for discrete outcomes even with parametric restrictions. However, relevant instruments can enable more informative bounds than no-assumption bounds.

1.3.1 Generalized Instrumental Variable Model

Identification analysis is provided under the generalized instrumental variable (GIV) model restrictions in Chesher and Rosen (2017). Under restrictions 1-6

in their paper, the identified set for the structural function h and a collection of conditional distributions $\mathcal{G}_{U|Z} \equiv \{G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$ is characterized. The following restrictions satisfy all the GIV model restrictions.

Assumption 1. *Y and U are random scalars and X and Z are random vectors defined on a probability space $(\Omega, \mathcal{L}, \mathbb{P})$, endowed with the Borel sets on Ω.*

Assumption 2. *The support of Y is a subset of all non-negative integers $\mathcal{R}_Y \equiv \{0, 1, 2, \cdots\}$ and the support of (X, Z) is a subset of a finite dimensional Euclidean space. A collection of conditional distributions $\mathcal{F}_{YX|Z} \equiv \{F_{YX|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$ is identified by the sampling process where $F_{YX|Z}(\mathcal{T}|z) \equiv \mathbb{P}[(Y, X) \in \mathcal{T}|z]$ for all $\mathcal{T} \subseteq \mathcal{R}_{YX}$.*

Assumption 3. *U is uniformly distributed on the unit interval $[0, 1]$ and $G_{U|Z}(\cdot|z) = G_U(\cdot)$ for all $z \in \mathcal{R}_Z$ where for all $\mathcal{S} \subseteq [0, 1]$, $G_U(\mathcal{S}) \equiv P[U \in \mathcal{S}]$.*

Restriction 1 defines the probability space of random variables Y, X, U, and Z. Restriction 2 restricts the support of observable variables and requires identification of the join conditional distribution of Y and X given Z. Restriction 3 restricts the distribution of U and requires Z to be independent of U. As $\mathcal{G}_{U|Z}$ is singleton and known, the object of identification is only the structural function h which is fully characterized by the threshold functions $\{p_y(x)\}_{y \in \mathcal{R}_Y, x \in \mathcal{R}_X}$.

A key element of identification is the $\mathcal{U}$-level set, $\mathcal{U}(Y, X; h) \equiv \{u \in \mathcal{R}_U : h(X, u) = Y\}$. Then under the model (1.9), this set is simply $\mathcal{U}(y, x; h) = [p_y(x), p_{y+1}(x)]$. Let $\mathcal{S}$ be a closed subset of $[0, 1]$. The containment functional of $\mathcal{U}(Y, X; h)$ is denoted by

$$C_h(\mathcal{S}|z) \equiv \mathbb{P}[\mathcal{U}(Y, X; h) \subseteq \mathcal{S}|z]. \tag{1.10}$$

Let $\mathcal{H}^*$ denote the identified set of the structural function h and $\mathsf{F}(\mathcal{A})$ be the collection of all closed subsets of a set $\mathcal{A}$. Then Corollary 1 provides the sharp characterization of the identified set.

Corollary 1. *Under the model (1.9) and Restriction 1-3, the sharp identified set of the structural function h is defined as*

$$\mathcal{H}^* \equiv \{h : \forall \mathcal{S} \in \mathsf{F}([0, 1]), \ C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}), \ a.e \ z \in \mathcal{R}_Z\}.$$

1.3.2 Core determining test sets

The number of elements in $\mathsf{F}([0,1])$ is infinite and thus implementation of the characterisation in Corollary 1 is infeasible. To find an implementable characterization, a notion of core determining classes is employed as suggested in Galichon and Henry (2011). Under the model restrictions, a collection of core determining test sets (CDTS) is defined as follows.

Definition 1 (CDTS). *A subcollection of* $\mathsf{F}([0,1])$, Q_h, *is a collection of CDTS if for almost every* $z \in \mathcal{R}_Z$,

$$C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}), \quad \forall \mathcal{S} \in \mathsf{Q}_h \tag{1.11}$$

and (1.11) *implies the same inequality also holds for every* $\mathcal{S} \in \mathsf{F}([0,1])$.

Finding the smallest collection of CDTS is beneficial to reduce computational burden for identification of $\mathcal{H}^*$. Let U_h denote the support of $\mathcal{U}(Y, X; h)$.

$$\mathsf{U}_h \equiv \{[0, p_1(x)], \ [p_1(x), p_2(x)], \ \cdots, \ [p_y(x), p_{y+1}(x)], \ \cdots \ : \ x \in \mathcal{R}_X\}$$

Theorem 3 in Chesher and Rosen (2017) (TH3 henceforth) suggests a collection of *all connected unions* of elements in U_h (except $[0,1]$) as the collection of CDTS. Let $\tilde{\mathsf{Q}}_h$ be a collection of all connected unions of elements of U_h. Now I focus on the cases where X is discrete.[6] Then $\tilde{\mathsf{Q}}_h$ consists of multiple sets of intervals so that $\tilde{\mathsf{Q}}_h = \mathsf{U}_h \cup \mathsf{Q}_h \cup \mathsf{W}_h$ where

$$\mathsf{Q}_h \equiv \{[0, p_y(x)], [p_y(x), 1] : y \in \mathcal{R}_Y \backslash \{0\}, \quad x \in \mathcal{R}_X\} \tag{1.12}$$

$$\mathsf{W}_h \equiv \{[p_y(x), p_k(x')] : y, k \in \mathcal{R}_Y \backslash \{0\}, \ p_{y+1}(x) \leq p_k(x'), \ x, x' \in \mathcal{R}_X\}. \tag{1.13}$$

Further refinement from $\tilde{\mathsf{Q}}_h$ is achievable under a certain shape condition.

Condition 1. *For all* $y \in \mathcal{R}_Y$,

(a) Complete separation :

$$\max\{p_y(x_1), p_y(x_2), \cdots, p_y(x_K)\} \leq \min\{p_{y+1}(x_1), p_{y+1}(x_2), \cdots, p_{y+1}(x_K)\}$$

(b) Monotonicity :

$$p_y(x_1) \leq p_y(x_2) \leq \cdots \leq p_y(x_K) \quad or \quad p_y(x_K) \leq p_y(x_{K-1}) \leq \cdots \leq p_y(x_1)$$

---

[6]For continuous X, the number of elements in U_h is already uncountably many. Therefore, the number of CDTS is also uncountably many.

By exploiting Condition 1, I propose a refinement of $\tilde{\mathsf{Q}}_h$ in the following theorem.

Theorem 1. *Suppose that X is discrete. Under the model (1.9), Restriction 1-3 and Condition 1, Q_h is a collection of core determining test sets.*

This theorem is also applicable for bounded ordered outcomes. Identification of the structural function is straightforward. By replacing $\mathsf{F}([0,1])$ in Corollary 1 with $\tilde{\mathsf{Q}}_h$, the sharp identified set for h is characterized. Under Condition 1, using Q_h rather than $\tilde{\mathsf{Q}}_h$ suffices to obtain the sharp identified set.

Corollary 2. *Given the joint distribution of (Y, X, Z), the identified set for the structural function h is characterized as follows.*

$$\mathcal{H}^* = \{h : \forall \mathcal{S} \in \tilde{\mathsf{Q}}_h, C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \ a.e \ z \in \mathcal{R}_Z\}.$$

If Condition 1 is satisfied,

$$\mathcal{H}^* = \{h : \forall \mathcal{S} \in \mathsf{Q}_h, C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \ a.e \ z \in \mathcal{R}_Z\}.$$

For computational feasibility, I only focus on a finite subset of $\mathcal{R}_Y$. Since $p_y(x)$ converges to 1 as $y \to \infty$, a large enough integer $\bar{y}$ at which $p_{\bar{y}}(x)$ for all x are very close to 1 can be found. The values greater than $\bar{y}$ are almost never realized. These values are not of practical importance so negligible. Given data, one can use the largest realization of Y for $\bar{y}$ in practice. Now $\bar{\mathcal{R}}_Y \equiv \{0, 1, 2, \cdots, \bar{y}\}$ is of interest. $\tilde{\mathsf{Q}}_h$ and Q_h corresponding to $\bar{\mathcal{R}}_Y$ are also redefined. Given $K = |\mathcal{R}_X|$, the numbers of elements of Q_h and $\tilde{\mathsf{Q}}_h$ are $2\bar{y}K$ and $2\bar{y}K + \frac{\bar{y}(\bar{y}-1)}{2}K^2$ respectively. Each element of $\tilde{\mathsf{Q}}_h$ and Q_h provides a conditional moment inequality. Let $\tilde{T}$ and $\bar{T}$ denote the number of moment inequalities from $\tilde{\mathsf{Q}}_h$ and Q_h respectively. The ratio of $\tilde{T}$ to $\bar{T}$ is $\frac{K(\bar{y}-1)}{4} + 1$, which explosively increases as K and $\bar{y}$ grow. Therefore, the computational gain achieved by using Q_h becomes greater when K and $\bar{y}$ are large.

Remark 2. Condition 1 is highly restrictive. It is only satisfied for a particular set of structural functions. Nonetheless, the use of Q_h could be still beneficial without Condition 1 if it provides a good approximation of the sharp identified set.

The use of Q_h naturally leads to faster computation at the cost of identifying power. Note that this cost is specific to a structural function h. Suppose

the threshold functions are generated from a parametric structure i.e.

$$p_y(x) = F(y, \lambda(x)), \quad \lambda(x) = \exp(\alpha + \beta x) \tag{1.14}$$

where F belongs to a known class of parametric cdfs. Then monotonicity is satisfied and complete separation means β is very close to zero. α and β are partially identified by conditional moment inequalities in Corollary 2 and Q_h is core determining for values of β close to 0. In applied studies using partial identification, evaluating whether the identified set of β includes 0 is a main concern. A conclusion about the impact of X on Y can be drawn if the identified set contains only positive or negative values. Q_h provides strongest criterion to evaluate the values of β around 0. In other words, when the outer region provided by Q_h contains 0, the sharp identified set also includes 0.

For the values of β far from 0, Q_h does not provide sharp identifying power. However, in that case, Y exhibits more dispersion and hence the number of moment inequalities relevant to identification of h is larger. As Y becomes less discrete, the sharp identified set shrinks and so does the outer region. Therefore, strong enough identifying power can be afforded by Q_h in such a case. The proximity of the outer region to the sharp identified set depends on the underlying data generating process. In practice, my suggestion is to use Q_h first for identification and add more intervals from $\tilde{\mathsf{Q}}_h$ if the size of the outer region is too large.

Remark 3. If the outer region delivered by Q_h is large, further identifying power can be afforded by additional intervals such as $[p_y(x), p_k(x)]$ where $y, k \in \mathcal{R}_Y$ and $y < k$. These intervals provide additional identifying power for the values of β far from 0. However, how much it would be refined is the question left to empirical exercises.

The identification result here is fully nonparametric. The value of the containment functional is determined by the ordering among threshold functions. Therefore all possible orderings need to be considered for identification. Given a particular set of threshold functions, its ordering gives upper and lower bounds for each of its elements. If all elements indeed lie between their bounds, the particular set is included in $\mathcal{H}^*$. However, as the supports of Y and X become richer, the number of admissible orderings increases explosively.[7] Therefore, appropriate shape restrictions or parametric restrictions

---

[7]The number of admissible orderings is $(K(\bar{y}-1))!/((\bar{y}-1)!)^K$ as shown in Chesher and Smolinski (2012).

might be imposed in practice so as to reduce the number of admissible orderings.[8] Note that additional structural assumptions such as a triangular system do not play a role to reduce the number of orderings but they may provide additional identifying power for the structural function h.

1.4 Numerical illustration of identified sets

Geometry of identified sets is investigated using probability distributions of (Y, X, Z) given data generating processes. To avoid dealing with the tremendous number of possible orderings, parametric restrictions (Poisson and negative binomial) are imposed. The model is still partially identifying even with parametric restrictions but identified sets on numerical examples tend to be small. Parametric restrictions allow threshold functions to be generated by a smaller number of structural parameters. Let $\mathcal{H}^*$ denote the approximation of the sharp identified set (I call $\mathcal{H}^*$ the identified set henceforth, not necessarily sharp) delivered by Q_h. Suppose that J is the number of structural parameters. Then the algorithm to compute the identified set is following. Let $LB(y, x, z; \theta)$ and $UB(y, x, z; \theta)$ denote lower and upper bounds of $p_y(x)$ given z and the parameter vector θ.

- Define fine grid points on $\mathbb{R}^r$. Let Θ denote the set of grid points. Then $\Theta \equiv \{\theta_1, \theta_2, \cdots, \theta_J\}$ where J is the number of grid points in Θ.

- Generate the thresholds $\{\{p_y(x; \theta_j)\}_{y=1}^{\bar{y}}\}_{x \in \mathcal{R}_X}$ using θ_j. Then the ordering l_j among the thresholds is given.

- Compute $LB(y, x, z; \theta_j)$ and $UB(y, x, z; \theta_j)$ for all y, x, and z using the given ordering l_j and Corollary 2.

- Check whether all the following moment inequalities are satisfied. If so, include θ_j in $\mathcal{H}^*$. Otherwise, $\theta_j \notin \mathcal{H}^*$.

$$\forall x, y, z, \quad \begin{array}{ll} p_y(x; \theta_j) - LB(y, x, z; \theta_j) & \geq 0 \\ UB(y, x, z; \theta_j) - p_y(x; \theta_j) & \geq 0 \end{array} \tag{1.15}$$

- Repeat the above steps for all $j = 1, \cdots, J$.

---

[8]Nonparametric shape restrictions can also reduce the number of possible orderings between threshold functions. Chesher and Smolinski (2012) investigates implications of such restrictions e.g. complete separation, monotonicity, single and twin peakedness.

This algorithm delivers identified sets on a number of data generating processes throughout this section. $\bar{y}$ is defined as follows.

$$\bar{y} \equiv \min\{y : p_y(x) > 1 - 10^{-5}, \ x \in \mathcal{R}_X\}$$

All the identified sets in this section are computed analytically using population distributions of (Y, X, Z). I elucidate how to compute the identified set using this algorithm in Appendix A.

1.4.1 Poisson restriction

I specify a data generating process (DGP) to explore the geometry of identified sets.

$$Z^* \sim N(0, 1), \quad \begin{bmatrix} \varepsilon \\ V \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$

$$X^* = \delta_1 + \delta_2 Z + V$$

Z^* is independent of ε and V. X and Z are generated by binary indicators such that

$$Z = 1[Z^* \geq 0], \quad X = 1[X^* \geq 0].$$

Unobserved heterogeneity ε is normalised to $U \equiv \Phi(\varepsilon)$ where $\Phi(\cdot)$ is the standard normal cdf. To generate a count outcome, threshold functions are generated by the Poisson cdf.

$$p_{y+1}(X) = \exp(-\exp(\alpha + \beta X)) \sum_{m=0}^{y} \frac{\exp(\alpha + \beta X)^m}{m!} \tag{1.16}$$

Then a function $g(\cdot|X)$ generates Y by taking U as an argument.

$$g(\tau|X) \equiv \inf\{y : p_{y+1}(X) \geq \tau\}, \quad Y = g(U|X)$$

For identification of α and β, the Poisson restriction is imposed. The ATE of X is defined as $\exp(\alpha + \beta) - \exp(\alpha)$.

This DGP is convenient to understand 'identification at infinity'. δ_1 and δ_2 control the prediction power of Z on X. The strength of instruments is pivotal for the size of the identified set. If Z is a perfect predictor for X, the identified set becomes a point as endogeneity of X disappears. I compute identified sets using varying degrees of instrumental strength. Let J be a positive real number. When $\delta_1 = -J$ and $\delta_2 = 2J$, $\mathbb{P}[X = z|Z = z] \to 1$ as $J \to \infty$. The strong and super strong IVs have the values of J equal to 2 and 4 respectively.

Figure 1.1: Point estimation and set identification results

(a) Point estimates **(b)** Identified sets
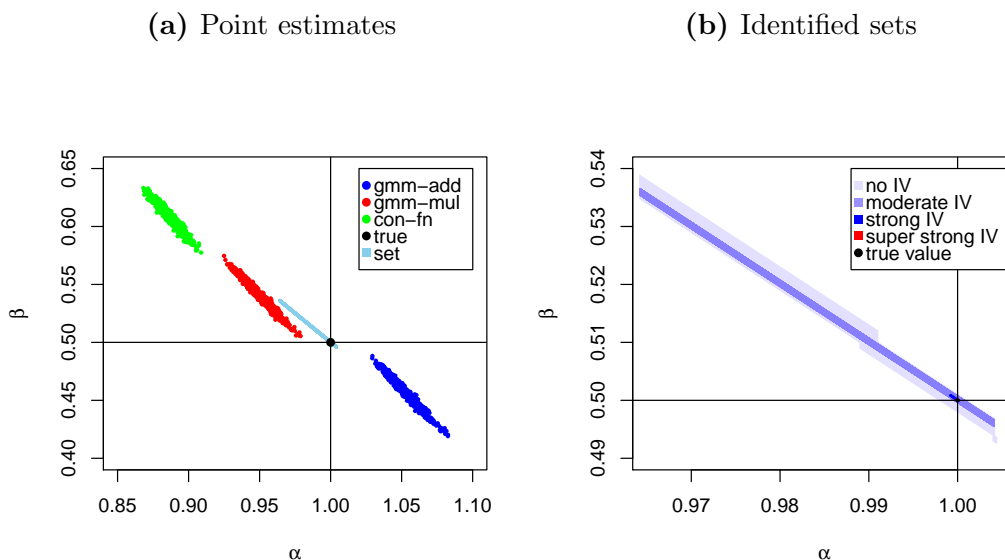
The moderate IV has $\delta_1 = 0$ and $\delta_2 = 1$ so that it only has one directional prediction power.

I first show that moment based models and the control function approach do not provide correct information about the true parameters under this DGP. The true values are used for the starting values of numerical optimization for moment based estimation.[9] Figure 1.1a shows the point estimates delivered by those methods with the moderate IV on $1,000$ Monte Carlo (MC) samples of length $n = 100,000$. 'gmm-add', 'gmm-mul', and 'con-fn' denote the additive and multiplicative moment based models, and the control function approach with the moderate instrument. They are all substantially far away from the true point. This is natural in the sense that those models are misspecified under the current DGP so the pseudo true values of point estimates differ from the true parameters in the DGP. The true parameters are only correctly backed out when the instrument is very strong, meaning that endogeneity is negligible. Furthermore, even with this large number of observations, those point estimates widely vary across MC samples.

On the contrary, the identified set with the moderate IV contains the true values and small enough to be informative. Figure 1.1b displays the identified sets associated with various instruments. Those sets are computed by exploiting the population distribution of observables. Considering the scale of the figure, those sets are very small. For the moderate IV, α and β lie

---

[9]Estimation results are robust to the choice of starting values if they are not very far from the true values.

Figure 1.2: Set identification under small support

(a) Outer region delivered by Q_h **(b)** Identified set with additional intervals
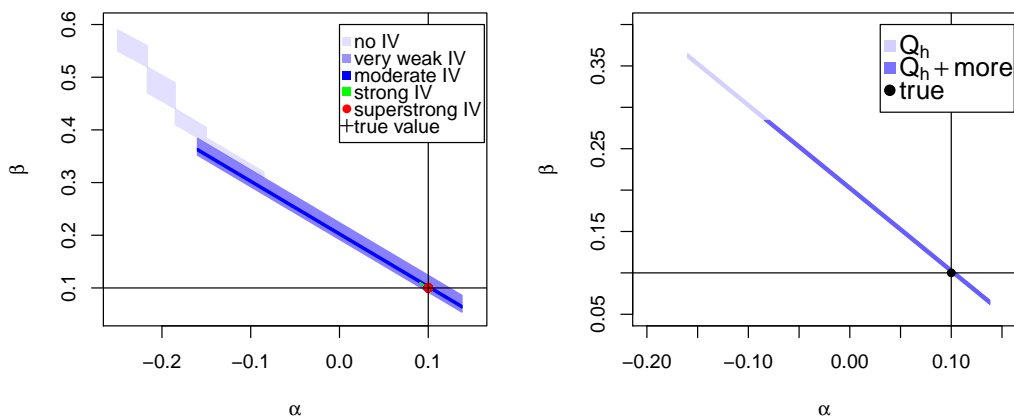
on $[0.497, 0.536]$ and $[0.965, 1.003]$ respectively. All the structural features of interest can be also computed from the identified set. The interval identified ATE is $[1.754, 1.860]$. For strong and super strong IVs, the identified sets are extremely small. The ATEs lie between $[1.763, 1.766]$ and $[1.7634, 1.7638]$ respectively. As $J \to \infty$, endogeneity in X disappears and the identified set converges to the true point.

The identified sets are small even if the IV is very weak. The source of strong identification power is the rich support of Y. Under the true parameter values, Y takes values from 0 to 17. In the case where $\alpha = 0.1$ and $\beta = 0.1$, the mean of Y is small and so is the variance. Y only takes values up to 9. Figure 1.2a shows the set identification results. Those sets are in general large unless instruments are very strong. However, even with Z having no correlation with X, the sign of the ATE is correctly identified. When $\delta_1 = \delta_2 = 0$, the identified sets for parameters and the ATE are

$$\alpha \in [-0.25, 0.135], \quad \beta \in [0.06, 0.59], \quad ATE \in [0.071, 0.626]$$

where the true ATE is 0.116.

Further refinement of the identified set in Figure 1.2a is afforded by additional intervals in $\tilde{\mathsf{Q}}_h$ excluded from Q_h. Reduction is mainly made on the area in which β is far from zero as predicted. Figure 1.2b shows the reduction when the instrument is moderate. These additional intervals are particularly useful if the outer region delivered by Q_h is large and connected and excludes

Figure 1.3: Sources of identifying power

(a) Identifying power of subsets of Q_h **(b)** Size variation of identified sets
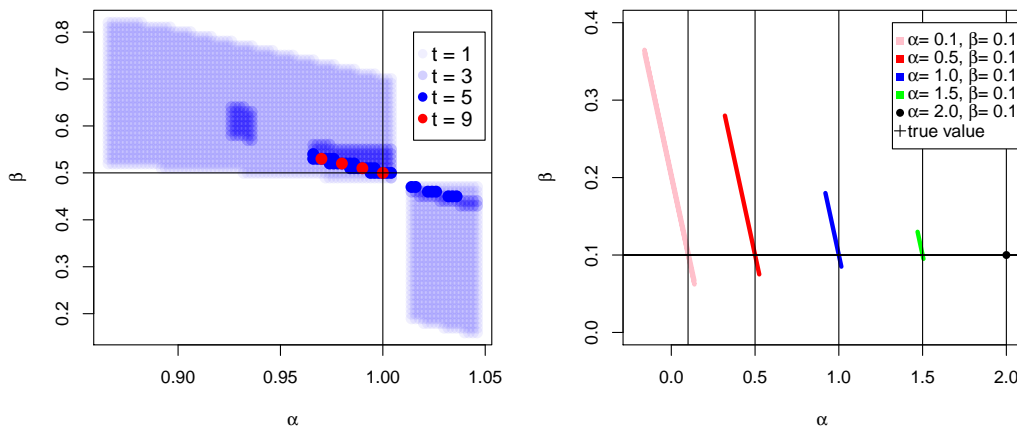
$\beta = 0$. One can expect a tighter set in such a case by exploiting more intervals in $\tilde{\mathsf{Q}}_h$. If the set is already tight enough, additional intervals would provide marginal identifying power. No evidence of size reduction is found when the support of Y is rich.

The GIV framework employed here does not necessarily require the rank condition if a model is partially identifying. Thus it is applicable in cases where an instrument is independent of U but has no prediction power for X. However, this identification power does not entirely come from this framework. In the example, X is positively correlated with U. Therefore, the observable joint distribution of Y, X and Z does not allow for negative values of β. If X is negatively correlated with U, then the identification power disappears. Let the correlation parameter γ be -0.5. Then the identified set contains the negative values of β. Therefore, in such a case, the identified set is not informative about the ATE.

One other interesting experiment is to evaluate identifying power of each intervals in Q_h. This experiment would answer the question : from which intervals the identifying power primarily comes? Under the current triangular system with $\alpha = 1, \beta = 0.5$, $\bar{y}$ is 17. For numerical identification, I use

$$\mathsf{Q}_h = \{[0, p_y(x)], [p_y(x), 1] : 1 \leq y \leq 17, X \in \{0, 1\}\}$$

which includes in total 68 intervals. Is it possible to deliver the same approximation of the identified set by a smaller number of intervals in

Q_h? Some evidence is shown in Figure 1.3a. A collection of intervals $\{[0, p_y(X)], [p_y(X), 1]\}_{y=1}^t$ is defined and the figure displays the outer regions delivered by different values of t. As t increases, the outer region converges to the identified set. Note that convergence is achieved at $t = 9$ given the density of grid points. This means the identifying power primarily comes from $\{[0, p_y(X)], [p_y(X), 1]\}_{y=1}^9$ and additional information provided by the other intervals is marginal.

Lastly, the richness of the support of Y is of great importance for the identifying power of the model. As $\mathcal{R}_Y$ becomes richer, the discreteness of Y decreases. Figure 1.3b shows the size variation of identified sets with respect to the richness of $\mathcal{R}_Y$ when the IV is moderate. In each DGP, the value of α varies. The larger α means the higher mean of Y. By equidispersion of the Poisson distriubution, the variance of Y also goes up with its mean. Therefore, the larger α is translated to the richer support of Y. The size of the identified set shrinks as α goes up. When $\alpha = 2$ ($E[Y] = 7.91$), the identified set becomes undistinguishable from a point. Therefore, the high dispersion in the count outcome we have, the smaller identified set is delivered.

1.4.2 Negative binomial restriction

When the threshold functions are generated by the negative binomial (NB) cdf, An additional shape parameter κ is involved. The identified set resides in a 3-dimensional space.

$$p_{y+1}(X) = \left(\frac{\kappa^{-1}}{\kappa^{-1} + \exp(\alpha + \beta X)} \right)^{\kappa^{-1}} \sum_{m=0}^{y} \frac{\Gamma(\kappa^{-1} + m)}{\Gamma(\kappa^{-1})\Gamma(m+1)} \left(\frac{\exp(\alpha + \beta X)}{\exp(\alpha + \beta X) + \kappa^{-1}} \right)^m \tag{1.17}$$

The conditional mean of Y given X is equal to that of the Poisson specification but the conditional variance differs. If α and β are identical to those of the Poisson distribution, the NB distribution converges to Poisson as κ goes to zero.

The true parameter values are $\alpha = 1, \beta = 0.5$ and $\kappa = 1$. Two instruments are employed, strong instrument ($J = 4$) and moderate ($\delta_1 = 0$, $\delta_2 = 1$). The 3-dimensional identified set with the moderate IV is computed and its convex hull is shown in Figure 1.4. The intersection of red lines in the figure indicates the true parameter values. The identified set is still small enough to be informative. The identified interval for each parameter is as follows.

$$\alpha \in [0.999, 1.001], \quad \beta \in [0.498, 0.502], \quad \kappa \in [0.999, 1.001]$$

Figure 1.4: The identified set under the NB structure
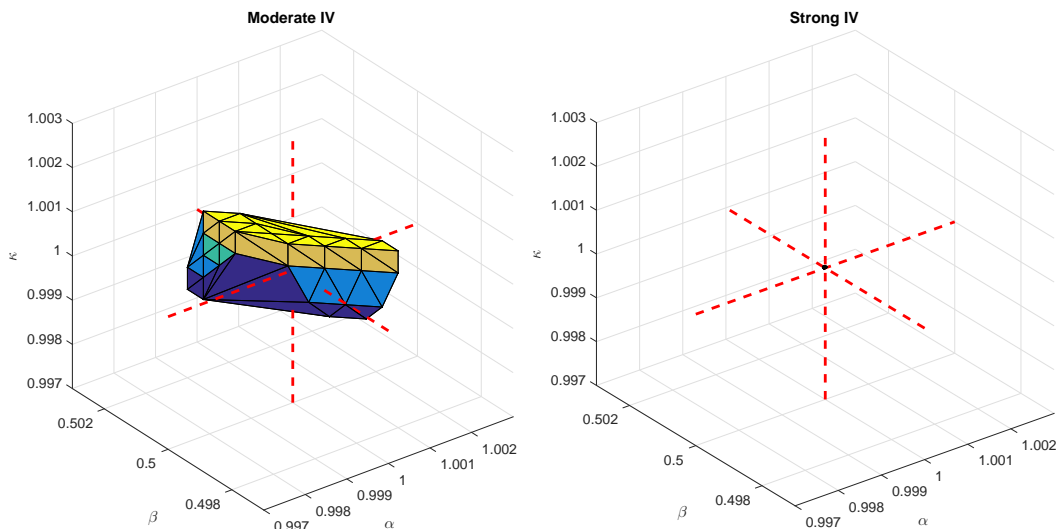
For the strong IV, the identified set is indistinguishable from a point given the scale of the figure. The interval identified α, β and κ are $[0.99998, 1.00002]$, $[0.49998, 0.50005]$ and $[0.99997, 1.00003]$ respectively.

Remark 4. If the true DGP is the NB structure, the Poisson restriction may deliver an empty set. That means there exists no parameter combination which generates the observed data distribution under the Poisson restriction. The Poisson restriction is misspecified in such a case. However, misspecified restrictions do not always deliver an empty set.

1.5 Estimation and Inference in Empirical Application

The proposed partial identification framework is easily implemented in a finite sample. Naive sample analogue estimators for upper and lower bounds of threshold functions are consistent. However, unlike the identification analysis, the population distribution of observable variables is left unknown. The sample analogue estimator is known to be biased in a finite sample as shown in Manski and Pepper (2009) even in the case where the support of Z is finite. The lower (upper) bounds tend to be upward (downward) biased and hence the bound estimates may be substantially tighter than the true bounds. In general the more points in $\mathcal{R}_Z$ or the smaller sample size, the larger magnitude of the finite bias. The finite sample bias results in serious problems. As shown in numerical examples, bounds can be very tight in count data models. Therefore,

the identification algorithm might deliver an empty set in a finite sample, even though the true identified set is perhaps non-empty.

To overcome this problem, a novel inference method introduced in Chernozhukov *et al.* (2013) is employed. Given the parameter grid Θ, the null and alternative hypotheses are

$$H_0 : \theta_j \in \mathcal{H}^* \quad v.s \quad H_1 : \theta_j \notin \mathcal{H}^* \tag{1.18}$$

for each grid point. Let $\hat{LB}(y, x, z; \theta_j)$ and $\hat{UB}(y, x, z; \theta_j)$ be sample analogue estimators of the lower and upper bounds of $p_y(x)$ given z respectively. Define sample moment functions

$$\begin{aligned} \hat{g}_L(y, x, z; \theta_j) &\equiv p_y(x; \theta_j) - \hat{LB}(y, x, z; \theta_j), \\ \hat{g}_U(y, x, z; \theta_j) &\equiv \hat{UB}(y, x, z; \theta_j) - p_y(x; \theta_j). \end{aligned} \tag{1.19}$$

Let

$$\hat{g}(\theta_j) \equiv \{\hat{g}_U(y, x, z; \theta_j), \hat{g}_U(y, x, z; \theta_j) : y \in \mathcal{R}_Y, x \in \mathcal{R}_X, z \in \mathcal{R}_Z\}. \tag{1.20}$$

Let L denote the number of elements of $\hat{g}(\theta_j)$. Then all the moment inequalities (1.15) are satisfied if and only if $\min \hat{g}(\theta_j) \geq 0$. Let $\hat{V}$ denote the sample covariance matrix of the moment vector $\hat{g}(\theta_j)$. Let $s = \sqrt{diag(\hat{V})/n}$ be the vector of standard errors of each sample moment.

For the critical value, draw a large number of a vector W from $N(0, \hat{\Sigma})$ where $\hat{\Sigma}$ is the correlation matrix corresponding to $\hat{V}$. Then a two step procedure produces the critical value $k_{1-\alpha}$ given the significance level $\alpha \geq 1/2$. The first step is to select moment conditions close to 0. Let $\bar{k}$ denote the γ_n-quantile of simulation draws $\max(W)$ where $\gamma_n \equiv 1 - 0.1/\log n$. Then the set of selected moments, $\hat{\mathcal{L}}$ is defined as follows.

$$\hat{\mathcal{L}} \equiv \left\{ l : g_l(\theta_j) \leq \min_{i \in \{1, \cdots, L\}} \{g_i(\theta_j) + \bar{k}s_i\} + 2\bar{k}s_l \right\} \tag{1.21}$$

Select $\tilde{W} \equiv (W_l : l \in \hat{\mathcal{L}})$. Then the critical value k_{1-a} is the α-quantile of $\max(\tilde{W})$ over simulation draws. Reject the null hypothesis only if

$$\min_{l \in \{1, \cdots, L\}} \{g_l(\theta_j) + k_{1-\alpha}s_l\} \leq 0. \tag{1.22}$$

1.5.1   Doctor visits and health insurance

CT13 demonstrates estimation and inference of count data IV models on cross-sectional data from the U.S. Medical Expenditure Panel Survey for 2003. The outcome is the annual number of doctor visits (DOCVIS) and the potentially endogenous variables are binary indicators of having private health insurance (PRIVATE). All the individuals in the sample are aged 65 or higher. The feature of interest is the causal effect of having additional health insurance, which supplements basic healthcare service through Medicare operated by the U.S. federal government, on demands for healthcare utilization.[10]

The sample contains 3,629 individuals aged 65-90. The maximum and sample mean of DOCVIS are 59 and 6.74 respectively. The sample variance is 45.56 which indicates overdispersion. No zero inflation is implied in the sense that only 10.9% of people in the sample never visited their doctors. The distribution of DOCVIS has a long right tail. Only less than 1% of the sample visited their doctors more than 40 times annually. 49.1% had private insurance.

PRIVATE is potentially correlated with unobserved factors in the sense that individuals with additional insurance chose to have it. They may have some personal reasons for supplemental insurance. Without controlling for endogeneity, regression analysis may suffer from selection bias. CT13 uses various count data IV methods to cope with endogeneity, such as the control function and moment-based approaches. They use income and the Social Security income share in total income (SSI) as instruments under the assumption that doctor visits of the old are not directly affected by income and SSI.

Table 1.1: Instrumental strength of INCOME

P[PRIVATE | income]	INCOME = 0	INCOME = 1
PRIVATE = 0	0.618	0.292
PRIVATE = 1	0.382	0.708

I employ the incomplete count data IV model (1.9) with parametric restrictions (Poisson and NB) and estimate the identified set for the coefficient on PRIVATE. Confidence regions of those set estimates are produced at various confidence levels. These results are compared to point estimates. A new variable INCOME is converted to a binary variable which takes value 1 if the income level is greater than its mean. I only use INCOME as an instrument.

---

[10]Medicare serves health insurance for all Americans aged 65 and older.

Figure 1.5: Set estimates and its 90% confidence region, and point estimates.
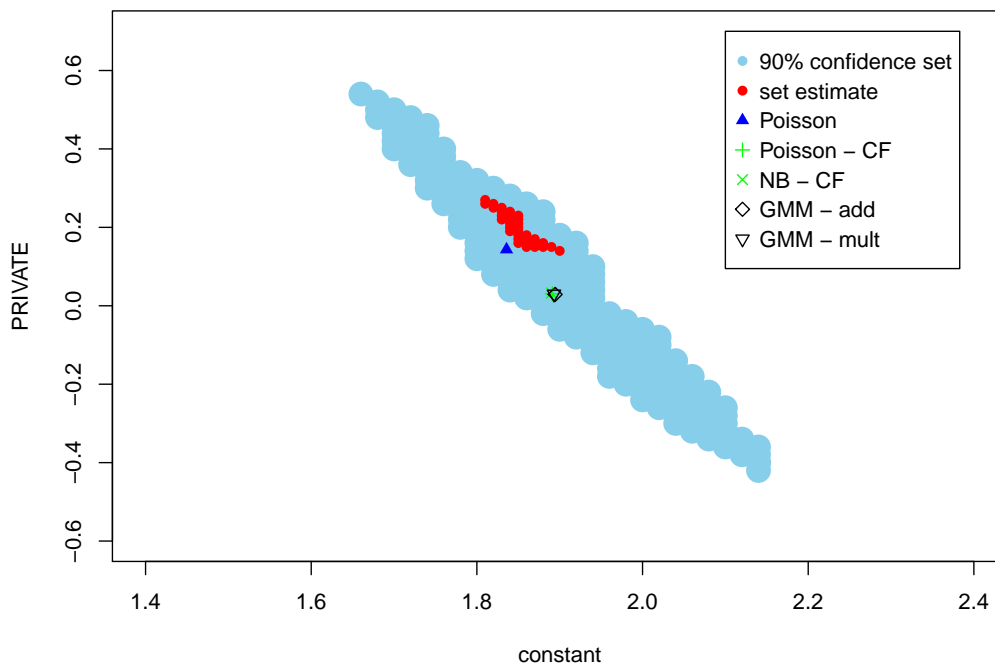
As shown in Table 1.1, income has substantial prediction power on the choice of supplemental insurance.

The model with the Poisson restriction delivers an empty set and an empty confidence region at any confidence level $\alpha \geq 1/2$. The empty set implies that the Poisson restriction is misspecified. On the contrary, the NB restriction provides with non-empty confidence regions for all $\alpha \geq 1/2$. I fix the shape parameter κ at 0.8 which is obtained from various parametric models as it is a nuisance parameter. Figure 1.5 displays set and point estimates of the coefficients on PRIVATE and the constant term. The set estimate (red dots) clearly differs from point estimates. The point estimates delivered by count IV methods are very close to each other, whereas the estimate from the Poisson regression with no control for endogeneity is much closer to the set estimate.

Table 1.2 shows the confidence interval of the coefficient on PRIVATE for each estimate. The set estimate of the coefficient on PRIVATE is computed at the significance level $\alpha = 1/2$ to correct the finite sample bias.[11] The set estimate is very different from point estimates all of which are close to 0

---

[11]This is called a half-median-unbiased estimator in Chernozhukov *et al.* (2013). With no correction, an empty set is delivered.

Table 1.2: Estimates of the coefficient on PRIVATE

	PRIVATE
Set estimate (NB)	(0.140, 0.270)
	[-0.420, 0.540]
Poisson	0.144
	[0.089, 0.198]
Poisson - Control function	0.030
	[-0.139, 0.200]
NB - Control function	0.033
	[-0.135, 0.200]
GMM - additive	0.029
	[-0.147, 0.205]
GMM - multiplicative	0.030
	[-0.151, 0.211]

Note: The box brackets indicate 90% confidence intervals. The confidence interval for the set estimate is the projection of the 90% joint confidence set.

except the Poisson model with no control for endogeneity. The treatment effect of additional health insurance implied by the set estimate is 15-31% increase in the number of doctor visits, meanwhile point estimates imply only 3% increase. As shown in the simulations studies, misspecification of the control function approach and moment based models cannot be ruled out.

As the sample size is not large enough, the size of confidence region at 90% level is too large to be informative. It is well known that the projection of an entire set estimate exhibits *projection conservatism* (see Kaido *et al.* (2017)). Hence the interpretation of the confidence region should be done with caution in the sense that individual parameters might have smaller confidence region.[12]

This empirical application is purely illustrative but it turns out to demonstrate that the set estimation framework is useful to read information from data. Set estimates and confidence regions can be reasonably tight to measure the effects of interest. It also suggests that misspecification can be captured by the framework. The empty set delivered by the Poisson restriction is a good example. Even though the sample size is not large and the instrument is not particularly strong in this application, implementation of the framework is straightforward and the set estimate provides useful information about the effect of private insurance on healthcare utilization.

---

[12]If a confidence interval is computed for each parameter, the confidence interval must be weakly smaller than the projection of the entire set. The degree of conservatism increases in line with the number of parameters in the model.

1.6 Conclusion

This chapter documents a unified approach to partial identification, set estimation and inference in incomplete count data IV models. The sharp characterization of the identified set is provided. The threshold crossing model specified here is flexible in the sense that it can nest most parametric count data models. The set estimation results in the empirical example deliver useful information about the effects of interest. It also implies misspecification of the Poisson restriction, which is not easily detectable in point identifying models. The example is done in an unfavourable setting in the sense that the sample size is not large and the instruments are not strong. In other cases where environments are more favourable, this approach is expected to deliver more informative results.

The bound analysis here indicates that practitioners should be cautioned when they use the commonly used count data methods under endogeneity. Point estimates delivered by those methods may be far from the true values of interest. A useful sensitivity analysis can be provided by the bound approach even if one still wishes to use those point estimates as shown in the empirical application of this chapter. Confidence regions of identified sets can be larger than desired in some unfavourable circumstances. Further developments in inference techniques such as Kaido *et al.* (2017) and Bugni *et al.* (2017) are expected to provide more powerful tools to make inference in such cases.

Chapter 2

Partial Identification in Competing Risks Models

2.1 Introduction

Applied researchers often observe durations terminated by one out of multiple competing causes. This arises for example in analysis of mortality data and unemployment durations. In the first type of data, each individual's age at death and the exact cause of death are reported. In the latter, an individual's unemployment spell is observed with a type of transition such as re-employment or exit from the labour force. Identification of cause-specific mortality or policy effects on unemployment durations until re-employment is not straightforward. Suppose an individual died from a heart attack. The person might have died from cancer or another disease if she did not have a heart attack but these counter-factual outcomes are censored. Focusing only on individuals who died from a heart attack without controlling for censoring in general delivers misleading information as they are likely to be more susceptible to that disease than the others.

Competing risks models (CRMs) are a useful tool to deal with this type of data. In these models, there exist multiple latent durations associated with causes but only the shortest duration is observed along with its cause. Objects of identification are the joint and marginal distributions of latent durations given covariates of interest. Marginal distributions are important for learning about responses of each duration to variations in covariate values. The joint distribution provides information about effects of changes in marginal distributions on overall survival probability. It has been known that the underlying distributions are unidentified without assuming a particular dependence struc-

ture among latent durations (Cox (1962), Tsiatis (1975)).[1] Peterson (1976) derives bounds without invoking restrictions on the dependence structure but these bounds tend to be too wide to be informative. The most common approach is to assume independence between durations but the assumption is often violated in applied studies.

The path-breaking paper by Honoré and Lleras-Muney (2006) (HL henceforth) develops a bounds approach to CRMs for discrete durations. The explicit modelling of the discreteness becomes an additional source of partial identification on top of the unknown dependence structure. One innovation of their approach was to allow for arbitrary dependence between durations. This approach places a strong parametric restriction under which covariate effects are only multiplicative. Another branch in the literature initiated by Zheng and Klein (1995) directly models the dependence structure between durations via widely used copula families. The copula approach can nonparametrically identify distributions of latent durations if the copula is known and durations are continuously measured. This approach provides richer information on covariate effects than parametric models. Identification of underlying distributions relies on numerical algorithms to solve a system of simultaneous nonlinear differential equations, which is computationally burdensome to solve. Moreover, asymptotic theory and inference methods do not exist except for some special cases.

In this chapter, I propose partially identifying CRMs for discretely measured or interval censored durations. Duration data are commonly measured on a discrete scale. For instance, unemployment spells are recorded in numbers of weeks, months or years. I exploit the discrete nature to derive nonparametric bounds on underlying distributions. Instead of restricting covariate effects, I model dependence directly using copula restrictions. Compared to the copula approach for continuously measured durations, the bounds require very little computational burden. Asymptotic properties of bound estimators can also be derived. I propose a valid bootstrap procedure which performs well in finite samples. The bounds approach is built upon copula restrictions which specify dependence between durations. There are no econometric tools to learn about true dependence. However, prior information on dependence can be obtained from auxiliary data, related literature or economic theory. If no information is available, the method provides useful sensitivity analysis for the varying degree of dependence.

---

[1]For any dependent joint distributions of durations, there exists an observationally equivalent joint distribution under which durations are mutually independent.

The bounds on marginal distributions are derived by solving a system of simultaneous nonlinear inequalities. In general, this type of system is extremely hard to solve and existence of solutions is not guaranteed. I show that, by rewriting the system of inequalities as a sequential system of equations, each bound is obtained as the unique solution of a simple nonlinear equation. Closed form expressions for the bounds are available under an independence assumption and otherwise any root finding algorithm can be employed to solve the bounds. Existing methods in the literature tend to focus on bivariate risks cases due to computational difficulties. The proposed approach in turn is not only extremely easy to implement but also computationally very efficient even for many hazards. In numerical examples, I demonstrate that this approach can provide tight bounds.

2.1.1 Related literature

In standard CRMs, $T^* \equiv (T_1^*, T_2^*, \cdots, T_J^*)$ is a random vector of latent durations associated with causes $j \in \mathcal{J} \equiv \{1, 2, \cdots, J\}$. T^* is assumed to be continuously distributed. The outcome vector Y^* is observed where

$$Y^* \equiv (Y_1^*, Y_2^*) \equiv (\min_{j \in \mathcal{J}} T_j^*, \arg\min_{j \in \mathcal{J}} T_j^*).$$

In the labour market, Y_1^* could be the spell of unemployment and Y_2^* the cause of leaving unemployment e.g. obtaining a job or exiting the labour force. In studying mortality, Y_1^* could be survival time and Y_2^* the cause of death. Tsiatis (1975) showed that CRMs are not identified without further restrictions. Since then numerous approaches have been devised to overcome this problem. Point identification of underlying distributions is only achieved by invoking strong structural restrictions which may sometimes be unjustifiable.

There is a large literature on identifiability of CRMs. Heckman and Honoré (1989) introduce conditions under which aspects of underlying distributions are identified within certain classes of models. This approach requires at least one of the regressors to affect underlying latent times differently and the regressors to have large supports. Abbring and Van den Berg (2003) show the large support assumption in Heckman and Honoré (1989) can be weakened in mixed proportional hazards CRMs. Lee (2006) identifies the joint survival function under a linear transformation regression model.[2] Lee and Lewbel (2013) develop identification of nonparametric accelerated failure time models with additive latent errors. Most papers in the literature assume latent times

---

[2]The joint distribution of latent failure times is not identified for all possible values of $(T_1^*, \cdots, T_J^*)$ so the marginal distributions of T_j is also left unidentified.

are continuously distributed. Han and Hausman (1990) model the discrete nature of duration data by employing an ordered choice structure. They showed that having more than two continuous covariates can identify the dependence structure of latent times under parametric distributional restrictions. Sueyoshi (1992) extends their results by accommodating time varying covariates. All the point identification results require the covariates X to be continuously distributed.

HL propose a model for discrete durations with restrictions on covariate effects. Suppose that T_j^* is a function of a binary covariate X. HL restricts the functional form such that

$$T_j^*(X) = \beta_j^X H_j, \quad X \in \{0, 1\} \tag{2.1}$$

where H_j are the baseline durations. In this case, X has a multiplicative effect on each duration. This model falls into a class of accelerated failure time models. The simple parametric restriction leads to set identification of β_j which captures the covariate effects on the j-th durations. They restrict the conditional distributions $F(t|X)$ and $F_j(t_j|X)$ such that

$$F(t|X = 1) = F(t_1/\beta_1, \cdots, t_J/\beta_J \mid X = 0), \quad F_j(t_j|X = 1) = F_j(t_j/\beta_j \mid X = 0).$$
$$\tag{2.2}$$

This means that the model only admits very specific forms of conditional distribution functions. While HL do not restrict the copula structure, they implicitly assume that the copula C is covariate-invariant such that $C(\cdot|X) = C(\cdot)$.

Another branch of the literature relies on the copula based approach to modelling the dependence structure between underlying hazards. No restrictions on covariate effects are imposed in this approach. Zheng and Klein (1995) were the first to apply this approach for bivariate CRMs and propose an estimator for marginal survival functions under the assumption that the copula is known. A closed form expression of this estimator is provided by Rivest and Wells (2001) when the known copula is of the Archimedean family. Carrière (1995) extends results of Zheng and Klein (1995) to cases with more than 2 risks. The marginal survival functions in such a case are identified by solving a system of simultaneous nonlinear differential equations numerically. This type of system is computationally hard to solve. Lo and Wilke (2010) suggest a risk pooling method with an Archimedean copula to reduce the computational burden.

The known copula assumption is restrictive but nonetheless invoked in many papers (see Chen (2010)). If the assumption is relaxed, some features can be partially identified. Fan and Liu (2018) partially identify the parameters of a linear quantile regression model under censoring by allowing the copula to vary within a certain class. Lo and Wilke (2016) set identify the sign of a covariate effect on marginal distributions. They make no parametric assumptions about the copula but requires it to be independent of covariates. Liu (2016) employs the Lévy copula to allow for time varying heterogeneity and simultaneous failure in the class of proportional hazards models.

The framework in this chapter is particularly attractive compared to existing methods because it is easy to implement in standard statistical packages, and computationally fast. A simple inference method is also available. Many papers in the literature do not provide methods to conduct inference. The trickiest feature of CRMs is that any assumptions on the dependence structure are untestable. Any identification results other than Peterson (1976) are entirely driven by model assumptions imposed. The partial identification approach is able to shed some light on understanding costs and benefits of additional identifying assumptions.

This chapter is structured as follows. Section 2 describes the proposed model. Identification results are also shown under parametric and nonparametric specifications. Section 3 demonstrates the numerical illustrations. Section 4 explains the estimation and inference procedures. Section 5 concludes. All technical proofs of main results are in the Appendix.

2.1.2 Notation

The notation in this chapter follows the convention in the literature. Upper and lower case letters A and a denote a random vector and its particular realization respectively. $\mathcal{R}_A$ denotes the support of A; $F(a|b)$ denotes the joint conditional distribution of $A = a$ given $B = b$; $F_j(a_j|b)$ denotes the marginal conditional distribution of $A_j = a_j$ given $B = b$ where A_j is the j-th element of A. The calligraphic font ($\mathcal{A}$) is used for sets and the sans serif font (A) is used for collections of sets.

2.2 Competing risks models

Durations are in general discretely measured or interval censored. Therefore, the researcher cannot observe the exact timing of an event. Define $\mathcal{T} \equiv \{0, 1, \cdots, M\}$ and $\mathcal{J} \equiv \{1, \cdots, J\}$. What can be observed is the mini-

mum of discretized latent durations $\{T_j\}_{j \in \mathcal{J}}$ such that

$$
\begin{aligned}
T_j \;=\; h_d(T_j^*) \;&=\; 0 && \text{if} && \tau_0 \le T_j^* < \tau_1 \\
&=\; 1 && \text{if} && \tau_1 \le T_j^* < \tau_2 \\
&=\; \cdots \\
&=\; m && \text{if} && \tau_m \le T_j^* < \tau_{m+1} \\
&=\; \cdots \\
&=\; M && \text{if} && \tau_M \le T_j^*.
\end{aligned}
\tag{2.3}
$$

If an event does not occur by τ_M, we cannot observe the failure. Thus the outcome may be top-coded. If times are measured in a discrete unit e.g. a week or a month, then $\tau_0 = 0$, $\tau_1 = 1$, $\cdots$, and $\tau_M = M$. The model can be either complete or incomplete depending on assumptions about availability of the exact cause. All the proofs of main results in this section are provided in Appendix II.

2.2.1 Complete model

Suppose the exact cause is observed. Then the outcome vector Y is observed by the researcher as follows.

$$
Y = (Y_1, Y_2) = (\min_{j \in \mathcal{J}} T_j, \arg\min_{j \in \mathcal{J}} T_j^*).
$$

Since T^* is continuous, it is natural to assume Y_2 is a singleton so that no tie is allowed. This model is complete in the sense that Y is pinned down given the values of latent durations T^*. The complete model setting is most widely used in the literature and is sensible in many cases. For instance, each unemployed individual can be only re-employed on either a full-time or part-time basis, but not both.

There may be a vector of covariates, X, that affect the distributions of T^*. Let $F_j(\cdot|X)$ denote the conditional marginal distributions of T_j^* and $F(\cdot|X)$ denote the conditional joint distribution of T^* given X. The object of identification is the set of conditional joint distributions

$$
\mathcal{F} \equiv \{F(\cdot|x) : x \in \mathcal{R}_X\}.
$$

Marginal distributions are identified by projections of $\mathcal{F}$. A set of mild assumptions are made.

Assumption 4. *Let $\delta_{mj}(x)$ denote $P[Y_1 = m, Y_2 = j | X = x]$. A set of condi-*

tional probabilities

$$\mathcal{L} \equiv \{\delta_{mj}(x) : m \in \mathcal{T}, j \in \mathcal{J}, x \in \mathcal{R}_X\}$$

is identified by the sampling process.

Assumption 5. *The conditional joint distribution of $(T_1^*, \cdots, T_J^*)$ given $x \in \mathcal{R}_X$ is such that the set $\{j : j = \arg\min_{j \in \mathcal{J}} T_j^*\}$ is a singleton with probability 1.*

Assumption 4 assures identification of conditional distributions of the outcome. Assumption 5 guarantees the uniqueness of Y_2. The conditional joint and marginal distribution functions are left unrestricted. They can be any proper distribution functions. No restrictions on $\mathcal{R}_X$ are imposed so that X can be either continuous or discrete. Point identification is not guaranteed under these assumptions. I use the identification results in Chesher and Rosen (2017) (CR17 henceforth) to derive the sharp bounds for underlying distributions. To utilise the framework in CR17, I define a structural function h as follows.

$$h(Y, T^*; h_d) = |Y_1 - h_d(T_{Y_2}^*)| + |Y_2 - \arg\min_{j \in \mathcal{J}} T_j^*|$$

Given the value of the outcome $Y = (m, j)$, a set of values of T^* which can generate the outcome is identified. The observed probability mass $\delta_{mj}(x)$ is equivalent to the likelihood attached to the set of values of T^*. Therefore, the following equality is satisfied.

$$\delta_{mj}(x) = P[\tau_m \leq T_j^* < \tau_{m+1} \wedge T_j^* < T_i^*, \forall i \neq j \mid X = x] \text{ for all } m \in \mathcal{T}, j \in \mathcal{J}. \tag{2.4}$$

The sharp identified set[3] of distributions $\mathcal{F}$ is characterised by a system of moment equalities.

Proposition 3. *Let $\mathcal{F}^*$ denote the sharp identified set of distributions $\mathcal{F}$. Under Assumption 4-5, $\mathcal{F}^*$ includes $F(x)$ for all $x \in \mathcal{R}_X$ such that (2.4) is satisfied for all $m \in \mathcal{T}$, $j \in \mathcal{J}$ given $F(x)$.*

In the absence of further assumptions, $\mathcal{F}$ is partially identified. However, computation of the identified set $\mathcal{F}^*$ using the characterisation is infeasible in the sense that one needs to search the whole distribution space to verify

---

[3]The sharp identified set consists of all conditional joint distributions of T^* which are able to produce the observational distribution $\mathcal{L}$ under the model assumptions. All distributions in $\mathcal{F}$ are observationally equivalent.

the conditional moment equalities. Instead, I first focus on identification of marginal distributions at the threshold values $\{\tau_m\}_{m=0}^{M}$. Define

$$p_{m,j}(x) \equiv F_j(\tau_m | x), \quad \mathcal{P}(x) \equiv \{p_{m,j}(x)\}_{m \in \mathcal{T}, j \in \mathcal{J}}.$$

It is convenient to use the concept of the copula in order to describe identification of marginal distributions.

Definition 2. *(copula) Let U denote $(U_1, \cdots, U_J)$ where U_j is uniformly distributed on the unit interval for all $j \in \mathcal{J}$. The joint distribution function of U is a copula $C(\cdot)$ such that*

$$C(u) \equiv \mathbb{P}[U_1 < u_1, \cdots, U_J < u_J].$$

By Sklar's Theorem,

$$F(T_1^*, \cdots, T_J^*) = C(F_1(T_1^*), \cdots, F_J(T_J^*)) \Leftrightarrow C(u) = F(F_1^{-1}(u_1), \cdots, F_J^{-1}(u_J))$$

and C is unique. $\mathcal{P}(x)$ is the set of values of margins given x at thresholds so is compatible with the definition of the copula. The joint distribution F is therefore decomposed into two components, the copula C and margins. Then by fixing C, bounds for marginal distributions can be derived. Now the object of identification is the set $\mathcal{P}(x)$ for all x. Further information on the marginal distribution functions at off-threshold values is not available. The joint distribution can be backed out given the copula and margins.

Bounds for $\mathcal{P}(x)$ are obtained via the following system of moment equalities and inequalities given restrictions on C. The following moment inequalities can be derived from (2.4).

$$\delta_{mj}(x) \leq P[\tau_m \leq T_j^* < \tau_{m+1} \wedge \tau_{m+1} \leq T_i^*, \forall i \neq j \mid X = x] \text{ for all } m \in \mathcal{T}, j \in \mathcal{J}$$
$$(2.5)$$

The right hand side (RHS) of (2.5) is the upper bound of the RHS of (2.4). Define

$$\gamma_m(x) \equiv P[Y_1 \leq m | x] = \sum_{j=1}^{J} \sum_{k=0}^{m} \delta_{kj}(x). \tag{2.6}$$

Then it is apparent that

$$1 - \gamma_m(x) = P[Y_1 > m | x] = P[\tau_{m+1} \leq T_j^*, \forall j \in \mathcal{J} \mid X = x] \text{ for all } m \in \mathcal{T}.$$
$$(2.7)$$

The RHSs of (2.5) and (2.7) can be written as functions of elements of $\mathcal{P}(x)$

given restrictions on the dependence structure between durations. Therefore, by solving the system of inequalities (2.5) and equalities (2.7), bounds for $\mathcal{P}(x)$ are derived.

2.2.1.1 Identification of marginal distributions given copula

I first illustrate a solution method for the system of nonlinear simultaneous equalities and inequalities. The following definitions are also useful to describe the solution method. (See Nelsen (2007) for further details of copula theory)

Definition 3. *(d-box) A d-box* $[\boldsymbol{a}, \boldsymbol{b}]$ *is the Cartesian product of d closed intervals such that*

$$[\boldsymbol{a}, \boldsymbol{b}] \equiv [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$$

where $\boldsymbol{a} = (a_1, \cdots, a_d)$, $\boldsymbol{b} = (b_1, \cdots, b_d)$ *and* $a_k \leq b_k$ *for all* $k \in \{1, \cdots, d\}$.

Definition 4. *(C-volume) For a given copula* $C(\cdot)$, *the C-volume* V_C *of a d-box is defined by*

$$V_C([\mathbf{a}, \mathbf{b}]) \equiv \sum_{v \in \mathcal{V}} sign(v) C(v)$$

where $\mathcal{V}$ *is a set of all vertices of the d-box* $[\boldsymbol{a}, \boldsymbol{b}]$ *and*

$$
\begin{aligned}
sign(v) = \quad & 1 \quad if \quad \sum_{k=1}^{d} 1[v_k = a_k] \quad is \ an \ even \ number \\
= \quad & -1 \quad if \quad \sum_{k=1}^{d} 1[v_k = a_k] \quad is \ an \ odd \ number.
\end{aligned}
$$

A d-box is a d-dimensional unit hypercube and a C-volume is the probability mass assigned to the d-box by the copula C. Figure 2.1 illustrates how to compute the C-volume in the 2-dimensional unit space given the independent copula. Let $\mathbf{a}$ and $\mathbf{b}$ be $(0.4, 0.3)$ and $(0.8, 0.7)$. Then the 4 vertices of the 2-box given $\mathbf{a}$ and $\mathbf{b}$ are

$$v_1 = (0.4, 0.3), \quad v_2 = (0.4, 0.7), \quad v_3 = (0.8, 0.3), \quad v_4 = (0.8, 0.7)$$

v_1 has two elements of $\mathbf{a}$, v_2 and v_3 have one and v_4 has none. Thus the C-volume is

$$V_C([a, b]) = C(v_4) - C(v_2) - C(v_3) + C(v_1).$$

as illustrated in Figure 2.1.

Figure 2.1: C-volume under the independent copula
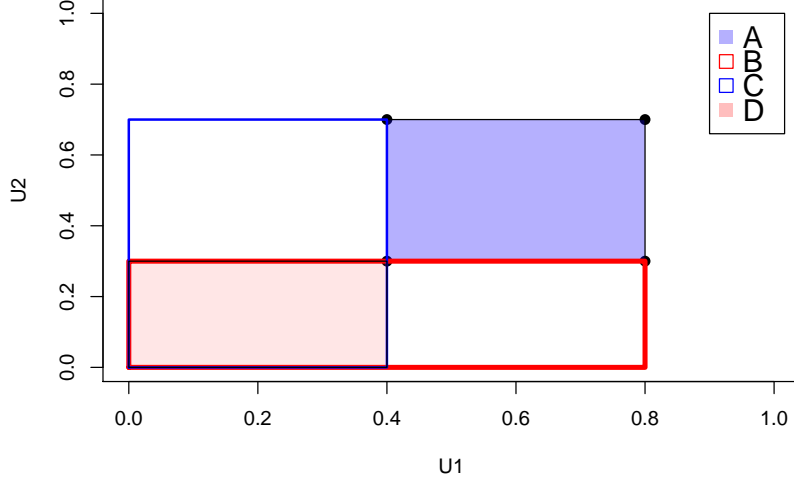
Note: The C-volume A is computed by subtracting $B(= C(v_2))$ and $C(= C(v_3))$ from the total area of the outer rectangle $(= C(v_4))$ and adding $D(= C(v_1))$.

Let $p^U_{m,j}(x)$ and $p^L_{m,j}(x)$ denote upper and lower bounds of $p_{m,j}(x)$. Define the following vectors.

$$\mathbf{a}_m(x) \equiv (p_{m,1}(x), \cdots, p_{m,J}(x)), \quad \mathbf{1}_{m,j}(x) \equiv (1, 1, \cdots, p_{m,j}(x), \cdots, 1, 1)$$

$\mathbf{1}_{m,j}(x)$ is a vector of length J in which the j-th element is $p_{m,j}(x)$ and 1 everywhere else. $\mathbf{1}$ is a one vector of length J. Now (2.5) and (2.7) can be rewritten as

$$\delta_{mj}(x) \leq V_C([\mathbf{a}_m(x), \mathbf{1}_{m+1,j}(x)]), \quad 1 - \gamma_m(x) = V_C([\mathbf{a}_{m+1,j}(x), \mathbf{1}]), \quad \forall m \in \mathcal{T}, \forall j \in \mathcal{J} \tag{2.8}$$

which define the bounds for $\mathcal{P}(x)$. It is apparent that $p_{0,j}(x) = 0$ for all j by construction. To fix ideas and for simpler exposition, I consider CRMs with only two hazards. Given the copula C, the following theorem provides the bounds.

Theorem 2. *Let Assumption 4-5 hold. For all $j \in \{1, 2\}$, $p^L_{1,j}(x) = \delta_{0j}(x)$. Given the lower bounds, unique values of $p^U_{m+1,1}(x)$ and $p^U_{m+1,2}(x)$ are obtained by solving*

$$\gamma_m(x) - p^L_{m+1,2}(x) = p_{m+1,1}(x) - C(p_{m+1,1}(x), p^L_{m+1,2}(x))$$

$$\gamma_m(x) - p^L_{m+1,1}(x) = p_{m+1,2}(x) - C(p^L_{m+1,1}(x), p_{m+1,2}(x))$$

for $p_{m+1,1}(x)$ and $p_{m+1,2}(x)$ respectively for all $m \geq 0$. Given $p^U_{m,1}(x)$ and $p^U_{m,2}(x)$, unique values of $p^L_{m+1,1}(x)$ and $p^L_{m+1,2}(x)$ are obtained by solving

$$\gamma_{m-1}(x) + \delta_{m1}(x) - p^U_{m,2}(x) = p_{m+1,1}(x) - C(p_{m+1,1}(x), p^U_{m,2}(x))$$

$$\gamma_{m-1}(x) + \delta_{m2}(x) - p^U_{m,1}(x) = p_{m+1,2}(x) - C(p^U_{m,1}(x), p_{m+1,2}(x))$$

for $p_{m+1,1}(x)$ and $p_{m+1,2}(x)$ respectively for $m \geq 1$.

Note that the upper and lower bounds of $p_{m+1,j}(x)$ are computed given the bounds of $p_{m,j}(x)$ for $m \geq 1$. Therefore, the bounds are sequentially derived from the smallest m to the largest. Given any copula C, the above theorem provides upper and lower bounds of the distribution functions at the threshold values. Each bound is derived from a single equation with one unknown. Thus computation is extremely simple. The RHSs of the equations are monotone in $p_{m+1,1}(x)$ or $p_{m+1,2}(x)$ so the solution is unique. Generalization of Theorem 2 to an arbitrary number of risks is provided in Appendix I.

Now consider the independence assumption.

Assumption 6. *(Independent copula) The joint distribution function of U is*

$$C(u) = \prod_{j=1}^{J} u_j.$$

This assumption implies that latent durations are independent of each other. Under the assumption, for any number of risks, closed form solutions can be found as shown in Lemma 1.

Lemma 1. *(Bounds under independence) Under Assumptions 4-6, the lower bounds are derived as follow. For all $j \in \mathcal{J}$,*

$$p^L_{1,j}(x) = \delta_{0j}(x), \quad p^L_{m+1,j}(x) = p^L_{m,j}(x) + \frac{\delta_{mj}(x)}{1 - \gamma_{m-1}(x)}(1 - p^L_{m,j}(x)) \quad \text{for } m \geq 1.$$

Given the lower bounds, the upper bounds are derived as follow.

$$p^U_{m,j}(x) = 1 - \frac{1 - \gamma_{m-1}(x)}{\prod_{i \neq j}^{J}(1 - p^L_{m,i}(x))}, \quad \forall\, m \geq 1,\; j \in \mathcal{J},\; x \in \mathcal{R}_X.$$

Now suppose no restrictions are imposed on the copula C so that any proper copulas are admitted. In such a case one can still derive bounds for

Figure 2.2: Illustration of moment inequalities
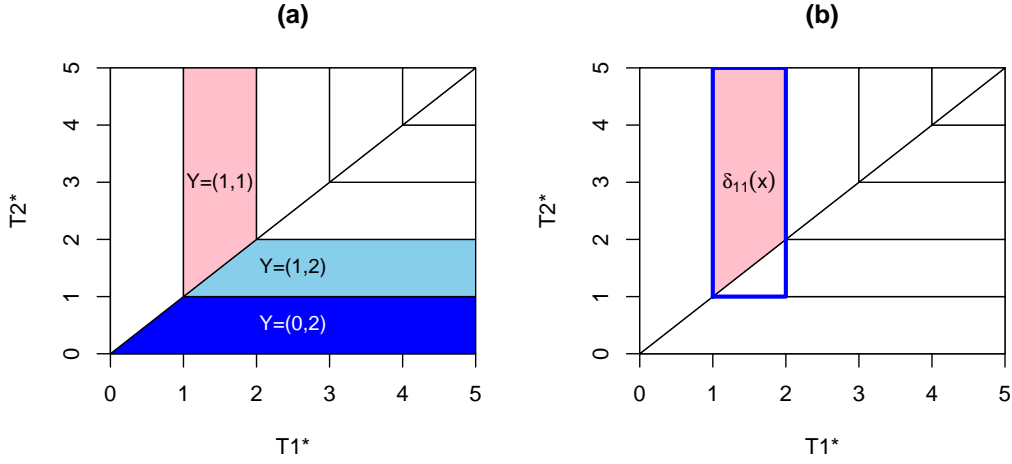
the margins, although they may not be very informative. Lemma 2 shows the worst case bounds when the researcher is not sure about the dependence structure among latent times.

Lemma 2. *(Worst case bounds) Under Assumption 4 and no other restriction on the copula C,*

$$p_{m+1,j}^{L}(x) = \delta_{mj}(x) + p_{m,j}^{L}(x) \quad and \quad p_{m+1,j}^{U}(x) = \sum_{j=1}^{J} \sum_{t=0}^{m} \delta_{tj}(x) \quad for \quad m \geq 0.$$

These bounds are sharp.

Illustration of identification : Suppose that $J = 2$, $M = 5$ and $\tau_0 = 0, \tau_1 = 1, \cdots, \tau_5 = 5$. Assume that T_1^* and T_2^* are smaller than 5 so that no top coding arises. Then, a set of pairs of (T_1^*, T_2^*) given the outcome $y = (m, j)$ is visualized in Figure 2.2 (a). The pink, blue, and light blue areas are linked to outcomes $(1, 1)$, $(1, 2)$ and $(0, 2)$ respectively. For the outcome $Y = (m, 1)$, (2.4) becomes

$$\delta_{m1} = P[\tau_m \leq T_1^* < T_2^* < \tau_{m+1}|x] + P[\tau_m \leq T_1^* < \tau_{m+1} \quad \wedge \quad \tau_{m+1} \leq T_2^*|x]. \tag{2.9}$$

If T_1^* and T_2^* are independent, (2.9) is simplified further.

$$\delta_{m1} = \int_{\tau_m}^{\tau_{m+1}} f_1(t|x)[p_{m+1,2}(x) - F_2(t|x)]dt + [p_{m+1,1}(x) - p_{m,1}(x)][1 - p_{m+1,2}(x)] \tag{2.10}$$

Figure 2.3: Illustration of moment equalities
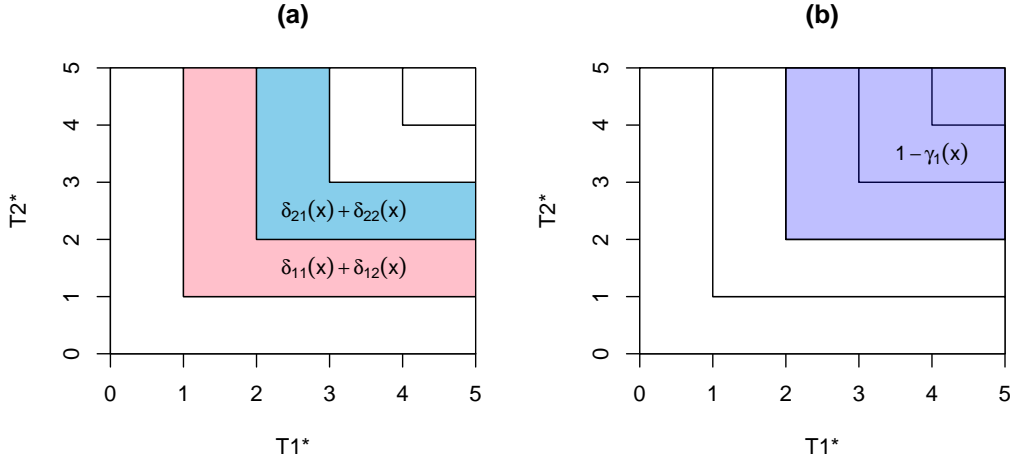
Without further restrictions, for $t \in [\tau_m, \tau_{m+1}]$ and $m \in \mathcal{T}$,

$$p_{m,2}(x) \leq F_2(t|x) \leq p_{m+1,2}(x), \quad \forall x \in \mathcal{R}_X. \tag{2.11}$$

This implies that (2.10) can be rewritten as

$$[p_{m+1,1}(x) - p_{m,1}(x)][1 - p_{m+1,2}(x)] \leq \delta_{m1}(x) \leq [p_{m+1,1}(x) - p_{m,1}(x)][1 - p_{m,2}(x)]. \tag{2.12}$$

Figure 2.2 (b) illustrates the RHS inequality for $Y = (1,1)$ in the sense that the probability $\delta_{11}(x)$ is naturally smaller than the area in the blue rectangle which is equal to $[p_{m+1,1}(x) - p_{m,1}(x)][1 - p_{m,2}(x)]$ under independence. The same logic applies to the outcome $Y = (m, 2)$ and thus bounds for $\mathcal{P}(x)$ are characterized by solving the system of the $4M$ number of quadratic polynomial inequalities. In general, these types of systems are hard to solve and sometimes bounds contain all values between 0 and 1. Therefore, I incorporate a set of equalities to make the system more tractable.

From (2.9), the m number of equalities are obtained by summing up $\delta_{m1}(x) + \delta_{m2}(x)$ as shown in Figure 2.3 (a). For all $m \in \mathcal{T}$,

$$\delta_{m1}(x) + \delta_{m2}(x) = P[\tau_m \leq T_1^* < \tau_{m+1} \ \wedge \ \tau_{m+1} \leq T_2^* \mid x]$$
$$+ P[\tau_m \leq T_2^* < \tau_{m+1} \ \wedge \ \tau_{m+1} \leq T_1^* \mid x] + P[\tau_m \leq T_1^*, T_2^* < \tau_{m+1} \mid x]. \tag{2.13}$$

The RHS of the equality is the probability mass assigned to each L-shape area

given the joint distribution. The set of equations (2.13) can be translated into

$$1 - \gamma_m(x) \;=\; P[\tau_{m+1} \le T_j^*, \; \forall j \in \mathcal{J} \mid x], \quad \forall m \in \mathcal{T} \tag{2.14}$$

by summing up equalities (2.13). As shown in Figure 2.3, $1 - \gamma_1(x)$ equals $\sum_{j=1}^2 \sum_{m=2}^4 \delta_{mj}(x)$. It is easy to show that (2.14) implies (2.13). Using (2.14) enables easy derivation of the nonparametric bounds. In the current example, (2.14) becomes

$$1 - \gamma_m(x) = (1 - p_{m+1,1}(x))(1 - p_{m+1,2}(x)), \tag{2.15}$$

which only has 2 unknowns.

Given $p_{0,j}(x) = 0$ for all x and j, it is straightforward that $\delta_{0j}(x) \le p_{1,j}(x)$ from (2.12) and hence the lower bound of $p_{1,j}(x)$ is $\delta_{0j}(x)$. From (2.15), $p_{m+1,j}(x)$ has the largest possible value when $p_{m+1,i}(x)$ is at its lower bound so that the upper bound of $p_{1,1}(x)$ is $1 - \frac{1-\gamma_0(x)}{1-\delta_{02}(x)}$. Now by substituting (2.15) into the RHS inequalities of (2.12),

$$p_{m,1}(x) + \frac{\delta_{m1}(x)}{1 - \gamma_{m-1}(x)}[1 - p_{m,1}(x)] \;\le\; p_{m+1,1}(x). \tag{2.16}$$

Noticing that $\frac{\delta_{m1}(x)}{1-\gamma_{m-1}(x)} < 1$, $p_{m+1,1}(x)$ has the lowest possible value at the lower bound of $p_{m,1}(x)$. Therefore, the bounds are sequentially derived using (2.15) and (2.16) from the smallest m to the largest. The LHS inequalities of (2.12) do not provide any additional information. Closed form expressions for those bounds turn out to be available under the independence assumption. Otherwise, given restrictions on the dependence structure, bounds can be derived by solving a system of equalities numerically.

2.2.2 Incomplete model

So far I have considered the cases in which the model is complete. The crucial assumption for completeness is Assumption 5. Define $\mathcal{M} \equiv \{j : \arg\min_{j \in \mathcal{J}} T_j\}$. T_j are discretized durations so the probability that $\mathcal{M}$ is singleton is non-zero. Now suppose that the researcher can only observe one cause among $j \in \mathcal{M}$. This could be the case when the death certificate of an individual is issued by a doctor. If multiple diseases are present when an individual dies, the doctor may choose any of them as a cause of the death using her subjective judgement.

Assumption 7. *Y_2 is a singleton such that $Y_2 \in \mathcal{M}$. If $\mathcal{M}$ is non-singleton,*

Y_2 is randomly determined by some unknown rule.

Given this assumption, the model becomes incomplete in the sense that knowing the values of T^* does not pin down the value of the outcome. For instance, in the two risks example, suppose $T_1^*, T_2^* \in [\tau_m, \tau_{m+1})$ so that $T_1 = T_2 = m$. Then $\mathcal{M} = \{1, 2\}$ and hence the outcome Y can be either $(m, 1)$ or $(m, 2)$.

Under this incomplete model, the sharp identified set of the distribution functions is characterized by a set of moment inequalities and a set of moment equalities as shown in CR17. The structural function h is now

$$h(Y, T^*; h_d) = |Y_1 - h_d(T_{Y_2}^*)| + \mathbb{I}[Y_2 \notin \mathcal{M}].$$

Define the U-level set

$$\mathcal{U}(m, j; h) \equiv \left\{ T^* : T_j^* \in [\tau_m, \tau_{m+1}] \ \wedge \ \tau_m \leq T_i^* \text{ for all } i \neq j \ \right\}. \qquad (2.17)$$

This is the set of all values of T^* that can produce the outcome $y = (m, j)$.

Theorem 3. *Let $\mathcal{F}^*$ denote the sharp identified set of $\mathcal{F}$. Let $\mathcal{N}$ be a proper subset of $\mathcal{J}$. Under Assumption 4 and 7, $\mathcal{F}^*$ includes $F(x)$ for all $x \in \mathcal{R}_X$ such that a set of moment inequalities*

$$\sum_{j \in \mathcal{N}} \delta_{mj}(x) \leq P[T^* \in \bigcup_{j \in \mathcal{N}} \mathcal{U}(m, j; h) | \ X = x]$$

are satisfied for all $m \in \mathcal{T}$, $\mathcal{N} \subset \mathcal{J}$ and a set of moment equalities

$$P[Y_1 > m | x] = P[\tau_{m+1} \leq T_j^* \text{ for all } j \in \mathcal{J} \mid X = x]$$

are satisfied for all $m \in \mathcal{T}$ given $F(x)$.

This theorem implies that the bounds derived in Theorem 2 are identical to the sharp bounds of $\mathcal{P}(x)$ under the incomplete model. For more than two hazards, the generalized version of Theorem 2 (Theorem 5 in Appendix I) provides the outer region in which the sharp identified set is nested.

2.2.3 Additional identifying assumptions

2.2.3.1 Parametric assumptions

Under parametric specifications of marginal distributions, the sharp characterization in Proposition 3 directly delivers the set of parameter values which are

able to generate the outcome distribution. The following assumption allows a finite number of parameters to characterize the distributions of latent failure times.

Assumption 8. *The joint conditional distribution of T^* and the marginal distributions of T_j^* for all $j \in \mathcal{J}$ given $x \in \mathcal{R}_X$ belong to a known class of parametric distributions.*

Let $F(\cdot|x; \theta)$ and $f(T_1, \cdots, T_J|x; \theta)$ denote the distribution and density functions respectively where θ is a finite vector of parameters. Define the following function.

$$g_j(t_j^*|x; \theta) \equiv \int_{t_j^*}^{\infty} f(t, \cdots, t, t_j^*, t, \cdots, t|x; \theta) \, dt$$

Let Θ^* be the identified set of values of θ. Under the complete model (Assumptions 4-5, and 8), Θ^* is delivered by conditional moment equalities in Proposition 3 as follows.

$$\Theta^* \equiv \left\{ \theta : \delta_{mj}(x) = \int_{\tau_m}^{\tau_{m+1}} g_j(t_j^*|x; \theta) dt_j^*, \ \forall m \in \mathcal{T}, \forall j \in \mathcal{J}, \forall x \in \mathcal{R}_X \right\} \quad (2.18)$$

Under the incomplete model assumptions, the identified set Θ^* is characterized in a similar fashion but with the set of moment inequalities and equalities in Theorem 3.

Suppose θ^* is the true value that generates the data. There may be multiple values of θ that are observationally equivalent to θ^*. θ is set identified in such a case. If Θ^* is a singleton, the model is point identifying. It is also possible that Θ^* is empty. This indicates the parametric restriction is misspecified so there exists no admissible structure generating the outcome distribution.

2.2.3.2 Exclusion restriction

There might exist a covariate Z which only affects one particular latent duration. In such a case, the variation in Z provides further identifying power for the other durations unaffected by Z. Suppose there are only two risks ($J = 2$) and a binary covariate Z affects only the first risk. The outcome distribution varies with Z but the marginal distribution of the second risk is unaffected by Z. The intersection of the two different bounds for the second risk given $Z = 0$ and $Z = 1$ provides tighter bounds. The exclusion restriction can be justified in many cases. For instance, if a new effective treatment for certain

types of diseases is approved, then it is likely to lead to lower mortality risks from those diseases without affecting the others.

I abstract from additional covariates and focus on the two risks case for notational simplicity but generalization is straightforward. The formal assumption is as follows.

Assumption 9. *There exists a covariate Z such that for all $z \in \mathcal{R}_Z$, $F_1(\cdot|Z = z) = F_1(\cdot)$.*

This assumption means that the covariate Z only affects the distribution of T_2^*, not T_1^*. Now given the value of $z \in \mathcal{R}_Z$, the set of conditional probabilities $\mathcal{L}(z) = \{\delta_{mj}(z) : m \in \mathcal{T}, j \in \mathcal{J}\}$ is identified. The following theorem maps this set to the bounds on $p_{m+1,1} \equiv F_1(\tau_m)$ and the bounds on $p_{m+1,2}(z) \equiv F_2(\tau_m|z)$ for all m.

Theorem 4. *Let Assumption 4-5 and 9 hold. Given $\mathcal{L}(z)$ for all $z \in \mathcal{R}_Z$,*

$$p_{1,1}^L = \sup_{z \in \mathcal{R}_Z} \delta_{01}(z), \quad p_{1,2}^L(z) = \delta_{02}(z).$$

Given the lower bounds, $p_{m+1,1}^U(z)$ and $p_{m+1,2}^U(z)$ are the unique roots of

$$\gamma_m(x) - p_{m+1,2}^L(z) = p_{m+1,1}(z) - C(p_{m+1,1}(z), p_{m+1,2}^L(z))$$
$$\gamma_m(x) - p_{m+1,1}^L = p_{m+1,2}(z) - C(p_{m+1,1}^L, p_{m+1,2}(z))$$

and $p_{m+1,1}^U = \inf_{z \in \mathcal{R}_Z} p_{m+1,1}^U(z)$ for all $m \geq 0$. Given the upper bounds, $p_{m+1,1}^L(z)$ and $p_{m+1,2}^L(z)$ are the unique roots of

$$\gamma_{m-1}(z) + \delta_{m1}(z) - p_{m,2}^U(z) = p_{m+1,1}(z) - C(p_{m+1,1}(z), p_{m,2}^U(z))$$
$$\gamma_{m-1}(z) + \delta_{m2}(z) - p_{m,1}^U = p_{m+1,2}(z) - C(p_{m,1}^U, p_{m+1,2}(z))$$

and $p_{m+1,1}^L = \sup_{z \in \mathcal{R}_Z} p_{m+1,1}^L(z)$ for $m \geq 1$ respectively.

2.3 Numerical examples

This section illustrates identification results in numerical examples. I specify a data generating process (DGP) from which the bounds are computed. Consider a case in which there are two competing hazards which induce two latent durations. Durations T_1^* and T_2^* are generated from the joint log-normal distribution and are positively associated.

$$T_1^* = \exp(U_1^*), \quad T_2^* = \exp(U_2^*)$$

where

$$\begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 + \beta_1 X \\ \mu_2 + \beta_2 X \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right)$$

Values of structural parameters are set to

$$\mu_1 = 0.125, \ \mu_2 = 0, \ \beta_1 = 0.25, \ \beta_2 = 0.125$$

$$\sigma_{11} = 0.25^2, \ \sigma_{12} = \frac{0.25^2}{2}, \ \sigma_{22} = 1.25 \times 0.25^2.$$

T_1 and T_2 are discretized durations such that

$$T_j = m, \quad \text{if } \tau_m \leq T_j^* < \tau_{m+1}, \quad m \in \{0, \cdots, 20\} \equiv \mathcal{T}, \qquad (2.19)$$

where the threshold values are defined by 20 equally spaced grid points between 0.5 and 1.7. Durations are top-coded at 20. The observable outcome consists of the minimum of two discretized durations and its exact cause.

$$Y = (\min\{T_1, T_2\}, \arg\min_{1,2}\{T_1^*, T_2^*\}) \qquad (2.20)$$

X shifts both margins downwards.

Given parameter values in the DGP, population probability masses $\delta_{mj}(x)$ can be computed for all m, j and x. I first compute worst case bounds using Lemma 2. Under no restriction on the relationship between latent durations, wide bounds are delivered as shown in Figure 2.4. These bounds are unfortunately not informative because no conclusion can be made except the downward shift of T_2 for short durations. These bounds are what we can learn from data if no further credible restrictions are available. Any particular restriction on the copula delivers a subset of the worst case bounds. Note that the worst case bounds do not restrict the copula to be invariant to covariates. Therefore, the worst case bounds contain the cases that the copula changes from one extreme to another e.g. perfect positive association to perfect negative association.

There might be some cases in which the researcher has prior information about the dependence structure. For instance, in medical studies, evidence can be found that certain diseases are driven by common factors such as genetic factors, smoking, drinking and obesity. It is well known that smoking causes not only lung cancer but also cardiovascular disease. In such a case, latent durations associated with those two hazards are likely positively correlated. The independent copula neither captures correct distribution functions nor

Figure 2.4: Worst case bounds for the marginal distributions of T_1 and T_2
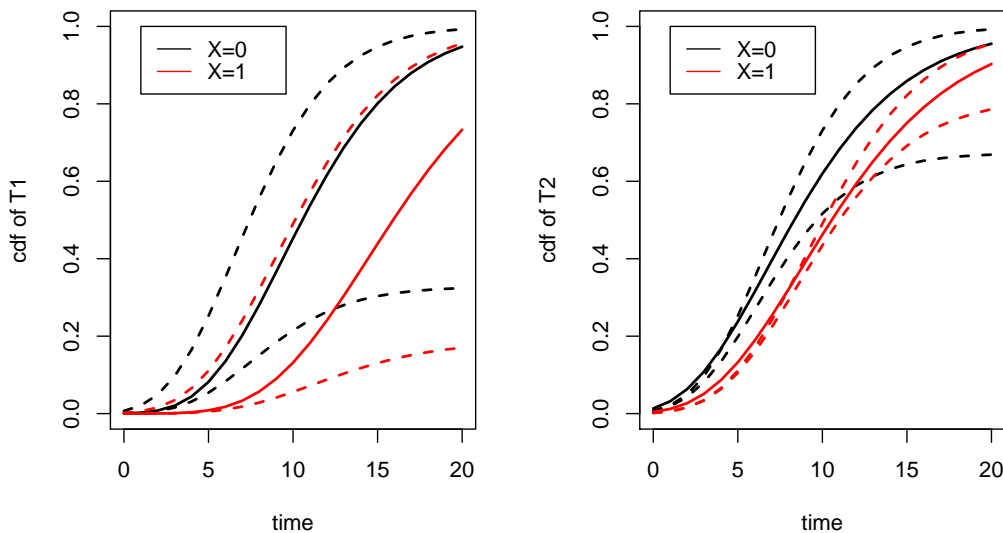
Note: Black and red solid lines are true marginal distribution functions conditional on $X = 0$ and $X = 1$ repectively. Dashed lines are upper and lower bounds for marginal distributions.

reasonably approximates them unless the correlation between the two is very weak.

Theorem 2 provides bounds given any copula. One can restrict the class of copulas to allow for dependence between durations. Gaussian and Archimedean copulas are good examples. Clayton, Frank and Gumbel copulas are the most widely used in the literature. They are Archimedean and are easy to deal with computationally since they only have one parameter which governs dependence between durations. The shape parameter θ is directly one-to-one related to commonly used dependence measures such as Kendall's τ and Spearman's ρ. Therefore, it is easy to compare the results from different families of copulas given the same dependence measure. Table 2.1 provides formulas and characteristics of those copula families for bivariate risks. Those three copulas have very different tail dependence.[4] The Frank copula exhibits no tail dependence. The Clayton and Gumbel copulas have opposite tail behaviours. The former has the lower tail dependence and the latter shows upper tail dependence.

---

[4]Tail dependence is the amount of dependence at the upper-right or lower-left tail of a multivariate distribution. One of many formal definitions of tail dependence can be found in Joe (1997).

**Table 2.1:** Archimedean copula families for bivariate risks models

Copula	$C(u_1, u_2)$	tail dependence
Clayton	$\left[u_1^{-\theta} + u_2^{-\theta} - 1\right]^{-1/\theta}$	lower tail
Frank	$-\frac{1}{\theta}\log\left[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$	no dependence
Gumbel	$\exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right\}$	upper tail

There is no way to learn about the true copula from data so it needs to be chosen by the researcher. Many papers using the copula based approach (Zheng and Klein (1995), Huang and Zhang (2008), Chen (2010) and Lo and Wilke (2010)) found that the choice of the copula family is much less important than the choice of the level of dependence (θ). They found that misspecified copula families result in slightly biased bounds when the degree of dependence is chosen correctly. It is also shown that a misspecified degree of dependence leads to severe bias in bounds. In the current example, restricting the copula to be of the Frank family brings very little identification power. The bounds in Figure 2.5 are computed by restricting the copula to be of the Frank family. The degree of dependence (measured in Kendall's τ) is only mildly restricted so that τ lies between -0.85 and 0.85. The bounds now capture the distributional shift of T_1 for very short durations.

Now suppose that auxiliary information suggests that two durations are moderately positively dependent. Then one can restrict the value of τ to lie in a certain range. By permitting $\tau \in [0, 0.5]$, much tighter bounds are calculated as shown in Figure 2.6. This exercise implies that strong identification power comes from restrictions on the degree of dependence.

If the copula is restricted to be of a certain family with a fixed parameter value, the marginal distributions are partially identified only due to the discreteness of the outcome. In the current example, the discreteness is not very severe in the sense that durations take values from 0 to 20. By fixing τ at the true level (0.295), very narrow bounds are obtained. Figure 2.7 shows bounds derived from three different copulas.

Considering that the true joint distribution is log-normal, those copulas perform well. The bounds from all copulas capture the true distribution for T_2^* almost perfectly. The fit for T_1^* is also fairly good. The Frank copula gives the best fit especially at the right tail. This results are reasonable in the sense that the log-normal distribution and the Frank copula have no tail dependence. Nonetheless, the results are not highly sensitive to the choice of

Figure 2.5: Bounds for the marginal distributions of T_1 and T_2 under Frank copula
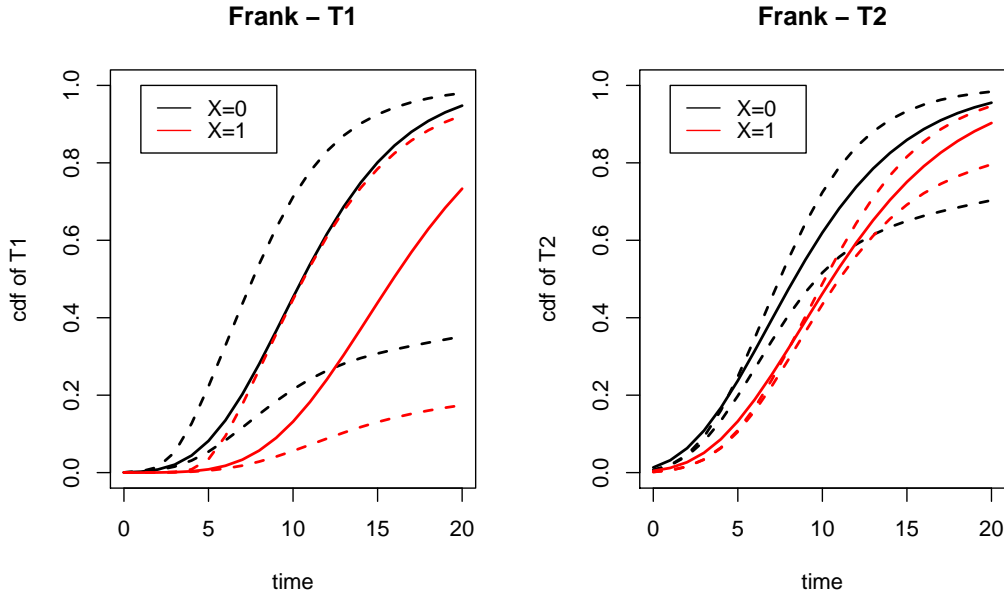
Note: Black and red solid lines are true marginal distribution functions conditional on $X = 0$ and $X = 1$ repectively. Dashed lines are upper and lower bounds for marginal distributions.

the copula family. If the association among risks is correctly specified, the marginal distributions are in general well approximated regardless of the form of copulas.

Given the marginal distributions, one can compute the mean of each latent duration. Comparing those means given also provides useful insights about improvements in each latent duration induced by changes in X. These average durations are understood as "expected cause-specific duration" which are counterfactual in the sense that they are computed as if there are no other risks. Given the fact that the survival function of T_j^* is $S_j(t|X) \equiv 1 - F_j(t|X)$, the mean of T_j^* is computed by integrating the survival function $S_j(t|X)$. Note that the full distribution functions are not always identified. The margins are identified up to $T = 20$ so they are truncated distributions. Therefore, I integrate the survival functions up to the truncation point.

Let $\mathbb{E}$ denote the expected value of a random variable from a truncated distribution such that

$$\mathbb{E}[T_j^*|X = x] \equiv \int_{\tau_0}^{\tau_M} S_j(t|X = x)dt. \tag{2.21}$$

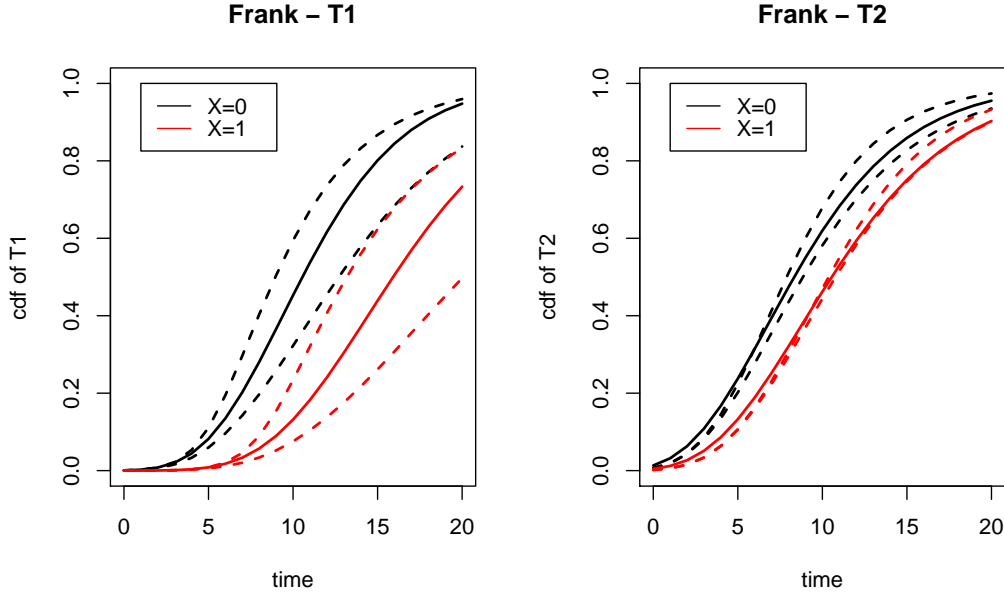**Figure 2.6:** Bounds under Frank copula and $\tau \in [0, 0.5]$

Note: Black and red solid lines are true marginal distribution functions conditional on $X = 0$ and $X = 1$ repectively. Dashed lines are upper and lower bounds for marginal distributions.

In the current example with the binary covariate X, growth in the mean is measured by the following ratio.

$$\eta_j = \frac{\mathbb{E}[T_j^*|X=1]}{\mathbb{E}[T_j^*|X=0]}, \quad j \in \{1, 2\}. \tag{2.22}$$

I define the ratio η_j as the "progress ratio" because it measures overall progress in each latent duration given the switch of X from 0 to 1. Given the bounds for the distribution functions, one can place bounds for η_j. For simplicity, I use the average of upper and lower bounds to compute the denominator of (2.22). Let $p_{m,j}^A(x)$ denote the mean of upper and lower bounds of $p_{m,j}(x)$. Then I compute the upper and lower bounds of η_j as follow.

$$\eta_j \in \left[\frac{\sum_{m=0}^{M-1}(\tau_{m+1} - \tau_m)p_{m+1,j}^L(1)}{\sum_{m=0}^{M-1}(\tau_{m+1} - \tau_m)p_{m+1,j}^A(0)}, \frac{\sum_{m=0}^{M-1}(\tau_{m+1} - \tau_m)p_{m+1,j}^U(1)}{\sum_{m=0}^{M-1}(\tau_{m+1} - \tau_m)p_{m+1,j}^A(0)} \right] \tag{2.23}$$

Figure 2.8 shows bounds for η_1 and η_2 given Clayton, Frank and Gumbel copulas with Kendall's $\tau \in [0.10, 0.80]$. The true values are contained within the bounds or very close to the boundary. Given the correct τ, all the copulas provide a very precise approximation of the overall progress ratios.

Figure 2.7: Bounds for distributions of T_1 and T_2 with Clayton, Frank and Gumbel copulas
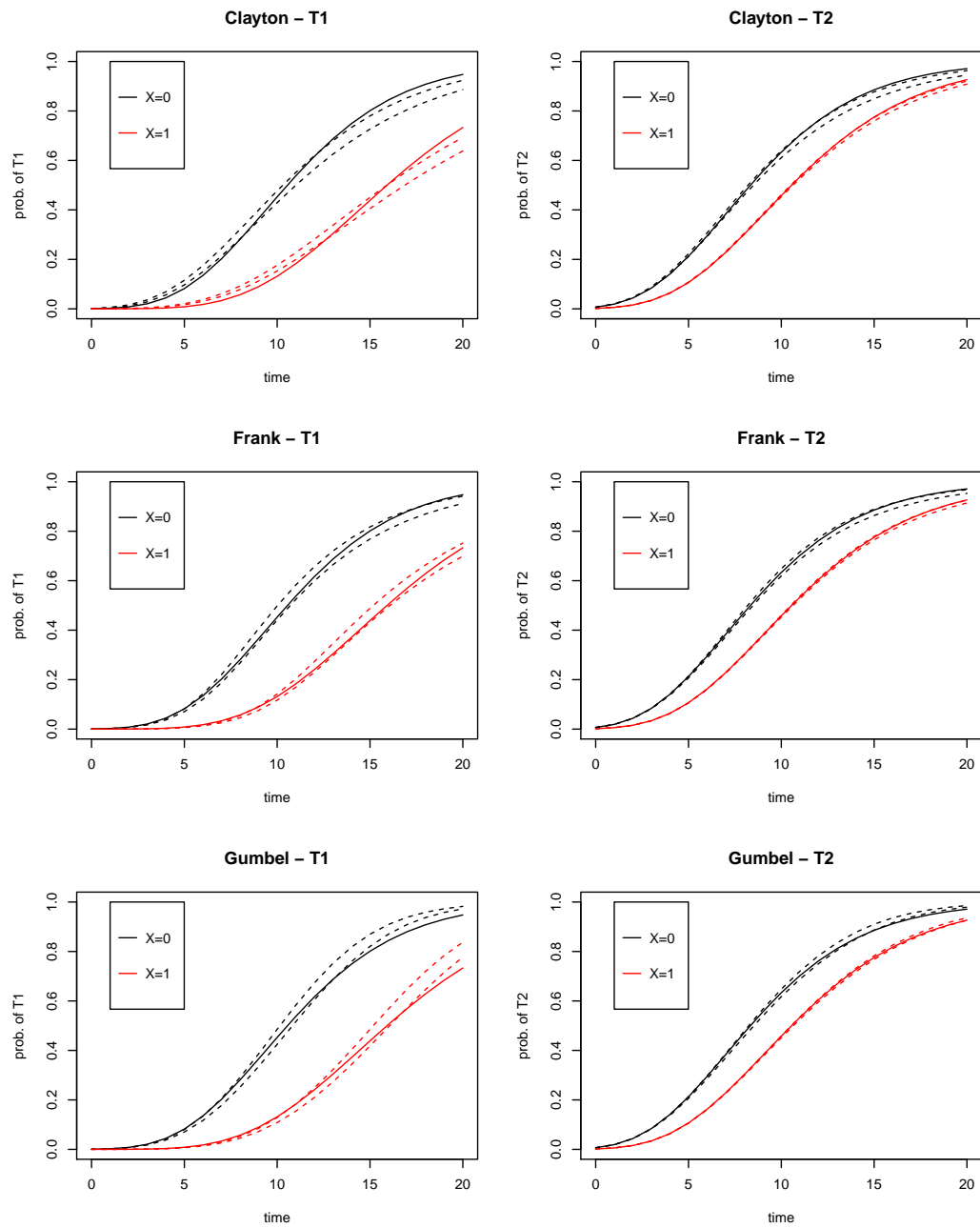
Due to the nature of CRMs, there are no tools to understand the dependence structure. If the copula varies with covariates, the worst-case bounds are all one can learn from the data. Therefore, additional restrictions have to be made to derive conclusive results if the worst case bounds are uninformative. The parametric copula families used in this section are convenient to allow for dependence between durations. These copulas provide useful insights on the

Figure 2.8: Bounds on progress in expected survival times
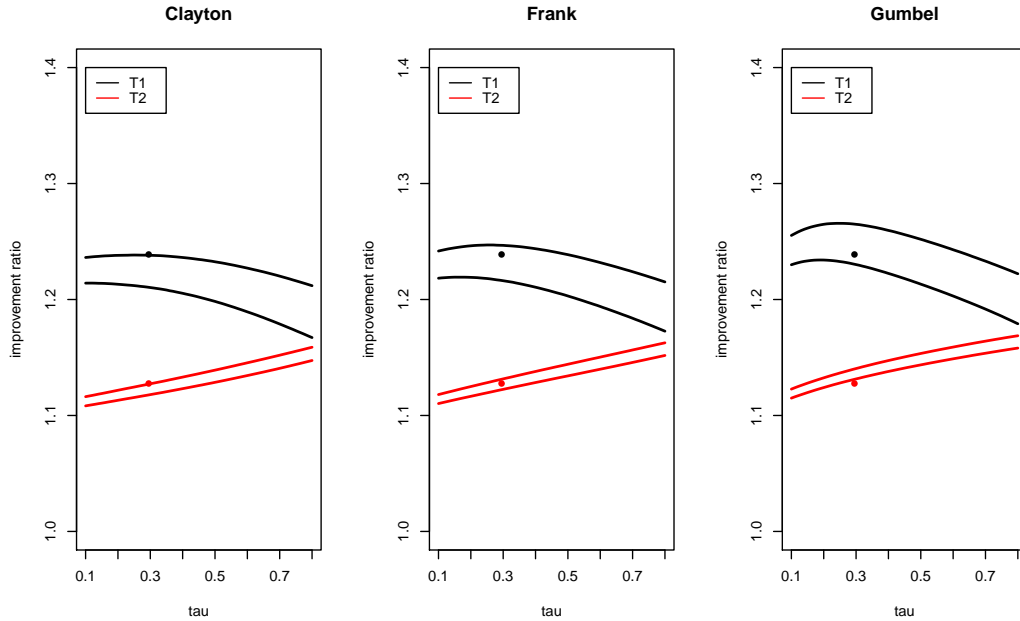
Note: Black and red solid lines are bounds on progress ratios for T_1 and T_2 respectively. Black and red dots are the true values of the progress in T_1 and T_2 given the DGP. The y-axis is the progress ratio and the x-axis is Kendall' τ from which the progress ratio is computed.

covariate effects given the plausible range of dependence. With no credible information on dependence, one can conduct sensitivity analysis and learn about the relationship between assumed dependence and covariate effects.

2.4 Estimation and Inference

The estimation and inference problem in nonparametric or semiparametric CRMs has not been widely studied. Zheng and Klein (1995) suggest the use of the jackknife variance estimator for their estimator. Asymptotic properties of their estimator are provided by Rivest and Wells (2001) only for bivariate risks with Archimedean copulas. Lo and Wilke (2010) provide bootstrap confidence bands without asymptotic theory for their estimator. Most other papers in the literature do not give results on inference.

There is a vast literature on confidence sets for partially identified parameters by moment inequalities. Horowitz and Manski (2000) propose a confidence interval that covers the entire identified set asymptotically with a prespecified probability in the context of missing data. Imbens and Manski (2004) develop a confidence interval for the parameter of interest, which is further extended

without superefficiency by Stoye (2009).[5] Chernozhukov *et al.* (2007), Andrews and Guggenberger (2009), and Romano and Shaikh (2010) derive confidence sets for identified sets using subsampling methods. Andrews and Han (2009) point out that the standard bootstrap fails to deliver valid inference in some particular cases and Bugni (2010) suggests a valid bootstrap procedure. Andrews and Guggenberger (2009) and Rosen (2008) consider "plug-in asymptotic" inference for elements of identified sets. Andrews and Soares (2010) and Andrews and Barwick (2012) propose moment selection methods to conduct asymptotically non-conservative inference. Kline and Tamer (2016) develop a Bayesian approach.

As Kline and Tamer (2016) point out, existing inference methods in the literature rely on "exhaustive search" over a parameter space using grid search to evaluate whether each grid point belongs to the confidence set. This naturally leads to a huge computational burden which explosively increases as the dimension of the parameter space grows and is therefore difficult to implement. The bound approach in this chapter is not reliant on grid search and hence a computationally tractable inference method can be drawn without exhaustive grid search.

In this section, I explain how to estimate the bounds using an i.i.d. sample $(Y_i, X_i)_{i=1}^{n}$ where $Y_i = (Y_{1i}, Y_{2i})$ and I introduce a simple bootstrap based inference method. The asymptotic properties of lower and upper bound estimators are provided. The bounds depend on sums of observed conditional probability masses $\delta_{mj}(x)$. One can find a consistent estimator of $\delta_{mj}(x)$ given data. For instance, if X has finite support, the consistent estimator of $\delta_{mj}(x)$ is simply

$$\hat{\delta}_{mj}(x) \equiv \frac{1}{n_x} \sum_{i=1}^{n} \mathbb{I}[Y_i = (m, j) \ \wedge \ X_i = x]$$

where $\mathbb{I}[\cdot]$ is an indicator function and $n_x \equiv \sum_{i=1}^{n} \mathbb{I}[X_i = x]$. This probability mass estimator converges to $\delta_{mj}(x)$ at the $\sqrt{n_x}$ rate. By the central limit theorem, under regularity conditions,

$$\sqrt{n_x}(\hat{\delta}_{mj}(x) - \delta_{mj}(x)) \to N(0, V_{mj}(x)) \ \text{ where } \ V_{mj}(x) = \delta_{mj}(x)(1 - \delta_{mj}(x)).$$

In the two hazards case ($J = 2$), for all $j \in \{1, 2\}$, it has been shown that $p_{1,j}^{L}(x) = \delta_{0j}(x)$. Therefore, its asymptotic distribution is trivial. The upper bound $p_{1,j}^{U}(x)$ is a function of $\delta_{0j}(x)$ and $p_{1,i}^{L}(x)$. Given that $p_{1,i}^{L}(x) = \delta_{0i}(x)$,

---

[5]Imbens and Manski (2004) implicitly assumes local superefficiency of estimation of a nuisance parameter which is the width of an identified set (interval).

define

$$p_{1,j}^U(x) = g_{1,j}^U(\delta_{0j}(x), \delta_{0i}(x)).$$

Given the differentiability of $g_{1,j}^U$, the asymptotic distribution of $p_{1,j}^U(x)$ is provided by the delta method. Likewise, define functions $g_{m,j}^L : [0,1] \times [0,1] \rightarrow [0,1]$ and $g_{m,j}^U : [0,1] \times [0,1] \rightarrow [0,1]$ for all m and j such that

$$p_{m+1,j}^U(x) = g_{m+1,j}^U(\gamma_m(x), p_{m+1,i}^L(x)), \quad p_{m+1,j}^L(x) = g_{m+1,j}^L(\gamma_{m-1}(x) + \delta_{mj}(x), p_{m,i}^U(x)).$$

Then asymptotic distributions of all bounds are sequentially derived by the delta method. Therefore, it is allowed to make pointwise inference on each bound.

However, implementation of these asymptotic distributions is practically impossible as the functions $g_{m,j}^L$ and $g_{m,j}^U$ in general do not have closed form expressions. Instead, confidence sets can be easily obtained by a bootstrap procedure. The standard bootstrap procedure yields asymptotically valid confidence sets. The bootstrap procedure to determine the $1 - \alpha$ confidence set is as follows.

- Step 1 : Draw a bootstrap sample (Y_i^b, X_i^b) from the original data with replacement where $b = 1, \cdots, B$. Compute the bootstrap probabilities $\{\hat{\delta}_{mj}^b(x)\}_{m \in \mathcal{T}, j \in \mathcal{J}, x \in \mathcal{R}_X}$.

- Step 2 : Compute the bounds using Theorem 2 for each bootstrap sample.

- Step 3 : Take the $(1 - \alpha)/2$ quantile of bootstrap upper bounds and the $\alpha/2$ quantile of bootstrap lower bounds.

Note that this inference procedure provides valid pointwise confidence bounds for each $p_{m,j}(x)$, not in a uniform sense. Uniform inference can be made using existing methods in the literature via grid search. This pointwise bootstrap inference does not rely on grid search so that the user can benefit from its computational efficiency. It requires little computational burden.

Monte Carlo simulations show that the proposed inference method works very well. I generate 2000 Monte Carlo samples to evaluate the finite sample performance of the bootstrap procedure. The sample size varies from 1000 to 10000. Two latent durations T_1^* and T_2^* are generated from the independent bivariate exponential distribution where the rate parameters are $\lambda_1 = 0.08$ and $\lambda_2 = 0.05$. The discretization process and the outcome Y are defined by (2.19) and (2.20). A coverage probability indicates the empirical probability

Table 2.2: Coverage probabilities of 95% bootstrap confidence bounds

Coverage prob.	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$
Overall	0.7615	0.8285	0.9290	0.9545
LB_1	0.9295	0.9490	0.9825	0.9885
UB_1	0.9255	0.9505	0.9755	0.9810
LB_1	0.9550	0.9660	0.9905	0.9955
UB_2	0.9310	0.9550	0.9780	0.9890

Note: Overall gives the probability that all upper and lower confidence bounds contain the true true distribution functions. LB_j denotes the probability that the lower bound for T_j is above the true distribution function. UB_j denotes the probability that the upper bound for T_j is below the true distribution function.

that confidence bounds computed from a Monte Carlo sample of size n contain the true distribution functions of T_1 and T_2. Table 2.2 displays coverage probabilities for sample size $n = 1000$, 2000, 5000, and 10000.

The coverage probability of each bound is about the nominal level or higher. The overall coverage probability is below the nominal level when sample size is 1000 but it goes up as sample size grows. It achieves satisfactory coverage when sample size is 5000. Note that those confidence bounds provide asymptotically correct coverage for the true upper and lower bounds of the distributions of latent durations. As the true bounds contain the true distribution functions, the confidence bounds deliver conservative coverage for the true distribution functions.

2.5 Conclusion

This chapter derives bounds on distribution functions of latent durations in competing risks models. The bounds approach is easy to implement and computationally more attractive than existing methods. Numerical examples show that those bounds can be narrow enough to be informative under appropriate assumptions. Simple estimation and inference methods are also introduced.

As competing risks models are fundamentally unidentified, the proposed method does not provide empirical tools to evaluate which restriction is more plausible. Therefore, care should be taken when one interprets results drawn from those restrictions. If risks are likely to be independent or there exist enough covariates so that risks are independent conditional on them, the closed form solutions under independence provides tight bounds. They can be also good approximations of true distribution functions if dependence between hazards is weak.

If one suspects that risks are likely dependent, the dependence structure can be modelled using some copula families. Prior information on dependence can be obtained from auxiliary data, related literature or economic theory. If no information is available, the method provides useful sensitivity analysis for effects of the varying degree of dependence on covariate effects.

Chapter 3

Applications of Partially Identifying Competing Risks Models

3.1 Introduction

CRMs are applied widely in a large number of fields including economics, biology, and medical science. Flinn and Heckman (1982), Katz and Meyer (1990), McCall (1996), Fallick and Ryu (2007), Farber and Valletta (2015) and Farber *et al.* (2015) study unemployment durations terminated by a number of transitions such as a transition to a full-time job or a part-time job, and exit from the labour force. Honoré and Lleras-Muney (2006) (HL) explores trends in mortality rates from cancer and CVD. Other interesting applications study PhD completion (Booth and Satchell (1995)), mortgage termination (Deng *et al.* (2000)), age at marriage or cohabitation (Berrington and Diamond (2000)), CEO exits (Gregory-Smith *et al.* (2009)), and bank failures (Wheelock and Wilson (2000)).

I conduct two empirical applications with the proposed bounds approach in Chapter 2. First, I revisit the mortality data studied in HL and evaluate trends in cancer mortality until 2010 by extending the original data set. Since President Nixon declared the "War on Cancer" in 1971, little reduction in cancer mortality has been found while cardiovascular disease (CVD) mortality has hugely declined. Little reduction in cancer mortality may be due to the huge fall in CVD in the sense that more people who might have died from CVD are exposed to risk of cancer. Empirical assessments that found little reduction in cancer mortality tend to assume independence between two diseases. However, cancer and CVD are likely to be dependent because common underlying fac-

tors drive both diseases.[1] Allowing for dependence between competing hazards may lead to entirely different conclusions.

The proposed bounds approach is closely related to HL and the copula based approach. HL's restriction implies that improvements in both types of mortality are homogeneous for all ages and that responses of marginal distributions to covariates are stochastically monotone. These implications are shown to be not compatible with patterns in mortality data which suggest the young (under age 65) and males have benefited more from reduction in cancer mortality. HL also only focuses on the bivariate hazards case, even though they have 4 different risk categories available in data. The reason is that their method is computationally very burdensome.[2]

The bounds approach shows heterogeneous improvements in cancer mortality across different ages, which cannot be captured under the parametric restriction used in HL. There has been greater reduction in mortality for the young. In addition, males, both white and black, experienced larger reduction in mortality than their female counterparts. Significant reduction in mortality is seen between 2000 and 2010. Allowing for dependence leads to larger estimates of the reduction in cancer mortality. The stronger the dependence assumed, the larger the magnitude found. These results imply the previous declaration that the war on cancer had failed may be false. Finally, subdividing cancers into two categories reveals different trends in mortality rates from lung cancer and all the other cancers.

In another empirical application, I estimate the effects of extended unemployment insurance (UI) benefits on unemployment spells during and after the great recession in the US using the data set from Farber and Valletta (2015). Farber *et al.* (2015) and Farber and Valletta (2015) assume that counter-factual durations until re-employment and exit from the labour force are independent but this assumption is likely to be violated as factors such as job search intensity and unobserved ability can drive both durations. It is more likely that they are negatively associated. If an individual has strong motivation for re-employment, she is more likely to have a short duration before re-employment but a long duration before exit from the labour force. I allow for negative dependence between the two and test the sensitivity of the results. The bounds

---

[1]Koene *et al.* (2016) lists many possible shared risk factors in cancer and CVD such as obesity, diabetes, tobacco smoking, diet and physical activities.

[2]Their method relies on a linear programming problem given a particular parameter value to check whether there exists a joint distribution of baseline hazards satisfying the model assumptions. It naturally involves grid search and therefore the computational burden explosively increases in line with the number of hazards and the support of covariates.

approach supports the original finding that extended UI benefits affected time to exit from the labour force but not time to re-employment during and after the Great Recession.

3.2 The war on cancer revisited

I revisit the mortality data studied by Honoré and Lleras-Muney (2006) (HL) and re-evaluate trends in cancer mortality. There had been no evidence of reduction in cancer mortality between 1970 and 2000 before HL found substantial reduction by allowing for dependence between cancer and cardiovascular disease (CVD). I calculate bounds on distributions of latent durations associated with CVD and cancer. The worst-case bounds are uninformative about trends in cancer mortality. The researcher has to place some restrictions on the dependence structure to derive conclusive results. Any restrictions on the dependence structure are untestable. However, the medical literature on the relationship between cancer and CVD provides guidance on which restrictions are plausible.

HL's approach has some limitations. Their restriction implies that reduction in mortality rates are homogeneous across ages. Cancer statistics suggest that cancer mortality has been reduced for young people but not for the old.[3] (SEER (2017)) Figure 3.1 shows age-adjusted mortality rates from CVD and cancer for under and over age 65. The mortality rate from CVD has significantly improved for both groups but cancer mortality has only declined for people aged under 65.

HL estimated largest reductions in cancer mortality for females than for males. However, both white and black males have experienced larger reductions than their female counterparts as shown in Figure 3.1. Furthermore, the gaps between estimated improvements under independence and dependence for females are much larger than for males as shown in Table 3.1. For white females, improvements in both mortality become comparable under dependence. The bounds approach proposed in Chapter 2 can shed light on whether HL's results are driven by their parametric restriction.

---

[3]It is natural that the younger generation has benefited more from cancer research and behavioural changes over time. As Cutler (2008) pointed out, progress in detection technologies had a substantial impact on survival from cancer. Cancer research also discovered various causes of cancer and people have modified their behaviour accordingly. Reduction in smoking is a good example.

Figure 3.1: U.S. Age adjusted mortality rates per 100, 000. See SEER (2017) for the details.
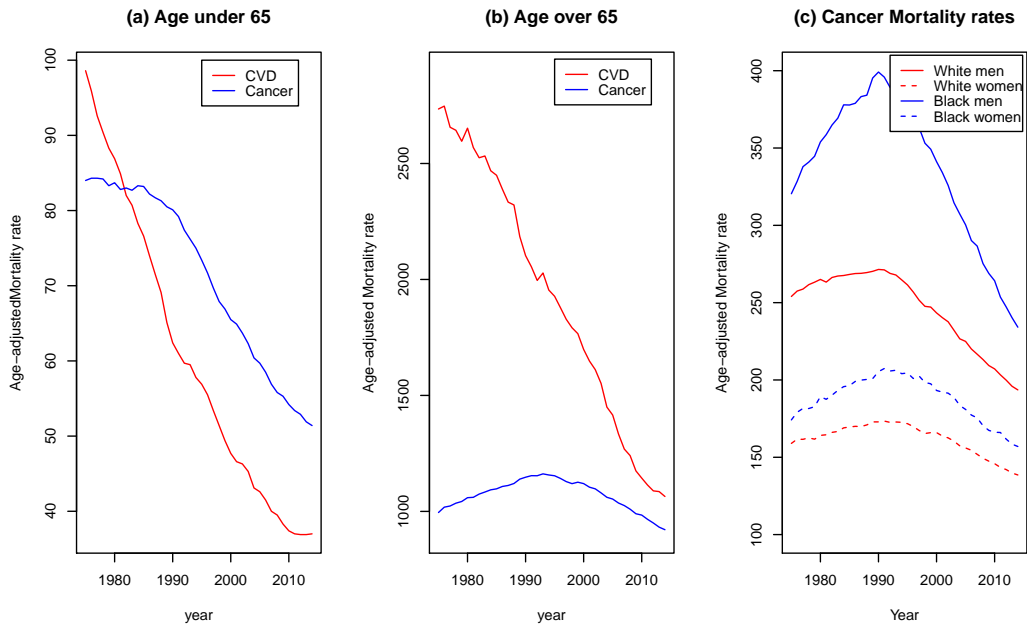
Table 3.1: Bounds for improvements in CVD and cancer mortality in HL

	CVD		cancer	
	indep.	dep.	indep.	dep.
White males	(1.392, 1.400)	(1.389, 1.391)	(1.059, 1.060)	(1.134, 1.153)
White females	(1.286, 1.291)	(1.236, 1.238)	(1.087, 1.093)	(1.201, 1.206)
Black males	(1.316, 1.320)	(1.334, 1.346)	(1.001, 1.029)	(1.072, 1.074)
Black females	(1.334, 1.346)	(1.334, 1.346)	(1.001, 1.029)	(1.160, 1.160)

Note: indep. and dep. mean under independence and dependence respectively. The numbers in parentheses are lower and upper bounds of the parameter β_j in (2.2), which measures improvements in mortality rates between 1970 and 2000.

3.2.1 Data

I use population data from the U.S. Census and Multiple Cause of Death data between 1980 and 2010 from which population probabilities of ages at death and causes of death are calculated. As the Census has been conducted decennially, I only use the mortality data from 1980, 1990, 2000 and 2010. I divide the causes into three categories: CVD, cancer and all other causes. The age-specific death counts induced by those causes are reported for four race-gender groups (black/white, male/female). Ages in the data are grouped from 0 to 100 so that the outcome variable naturally follows the discretization process (2.3). Details on the data sources are in HL.

Table 3.2: Fraction of deaths from each cause

cause	1980	1990	2000	2010	1980	1990	2000	2010
	White Male				White Female			
CVD	0.54	0.46	0.41	0.33	0.57	0.49	0.43	0.33
Cancer	0.23	0.27	0.27	0.27	0.22	0.24	0.23	0.23
(lung)	0.08	0.09	0.09	0.08	0.03	0.05	0.06	0.06
	Black Male				Black Female			
CVD	0.47	0.42	0.39	0.35	0.55	0.49	0.44	0.36
Cancer	0.25	0.29	0.28	0.27	0.21	0.23	0.23	0.24
(lung)	0.08	0.10	0.09	0.08	0.03	0.04	0.05	0.05

I do not censor the data at age 80. HL censored their data at age 80 because yearly mortality rates are only available up to age 80 for 1970. As a result, they focus on individuals aged between only 45 and 80 to compare mortality rates over time since 1970. However, for the later periods, yearly mortality rates are available up to age 100. For individuals aged between 80 and 100, trends in mortality are likely different from the people in HL's data. Therefore, I use all available mortality rates up to age 100 to capture rich information on mortality trends. This means that I only estimate mortality trends since 1980.[4] I restrict the sample to adults aged over 45. For people aged below 45, cancer and CVD occurrences are very rare. All results are conditional on survival to age 45.

Trends in fractions of deaths from CVD and cancer are quite opposite as shown in Table 3.2. For white males, the fraction of deaths from CVD kept decreasing from 54% in 1980 to 33% in 2010. However, the fraction of cancer jumped up from 23% to 27% over the same period. The similar pattern is found for other demographic groups. The overall fraction of CVD declined by 13-24% for all groups. The fraction of lung cancer is stable for males but increased over time for females. This shows a relatively large fall in CVD mortality, although it remains the largest mortality risk for all demographic groups.

In the following subsections, I estimate the bounds for the distribution functions of CVD and cancer with and without independence. I focus on two hazards (CVD and cancer) and extend the model for more hazards. Regard-

---

[4]This does not lose much information. HL and references therein found that cancer mortality was very stable between 1970 and 1980. Furthermore, Increasingly many people live over age 80 so it is important to take them into account. The life expectancy for Americans is 80 (CIA (2018)) which places the US 43rd in the world.

less of the number of hazards, theoretical results in Chapter 2 are applicable and computational costs remain low. Bounds tend to become slightly wider when the number of hazards increases but they still remain narrow in this application. I do not conduct inference in this application as I use population data. Given the number of observations, bound estimates are very precisely obtained.

3.2.2 Bounds under independence

I derive bounds for each race-gender group from 1980 to 2010. The distributional shifts captured by those bounds are understood as trends in mortality rates from CVD and cancer. These shifts cannot be interpreted as the effects of investments into cancer research. Many other factors such as environmental changes, changes in lifestyle may also lead to reduction in cancer mortality. Therefore, it is not feasible to separate pure effects of cancer research from other effects. What I estimate here is understood as overall improvements against CVD and cancer over time.

Figure 3.2-3.3 shows the bounds for each demographic group from 1980 to 2010. The mortality rate from CVD is constantly reduced over time for all groups. On the other hand, cancer mortality shows no reduction until 2000. It deteriorated for people over age 80. For males, small reduction is found between 1980-2000 for ages under 80. The blue solid lines deviate from the black solid lines but they converge to the black lines around age 85 for white men and around age 80 for black men. Significant reduction is shown between 2000 and 2010 for all ages. Both black and white women experienced disimprovements until 2000. Very small reduction is found between 2000-2010 for females. These findings are consistent with the patterns in Figure 3.1.

The results under independence show that trends in cancer mortality are heterogeneous across age groups. The overall improvements in mortality rates can be expressed as the "progress ratio" defined by (2.22). Table 3.3 displays the ratios of the counter-factual expected survival times in 1990-2010 to those in 1980. Those ratios provide insights on the overall mortality improvements since 1980. The mortality rates from CVD were improved by over 20% for men and by over 15% for women. Improvements in cancer mortality between 1980 and 2010 are around 3% and 5% for white and black men respectively. Women experienced less than 1% of improvements during the same period.

3.2.3 Bounds under dependence

It is reasonable to suspect that the independence assumption does not hold as cancer and CVD share common risk factors. In the medical literature,

Figure 3.2: Bounds on distribution functions of CVD
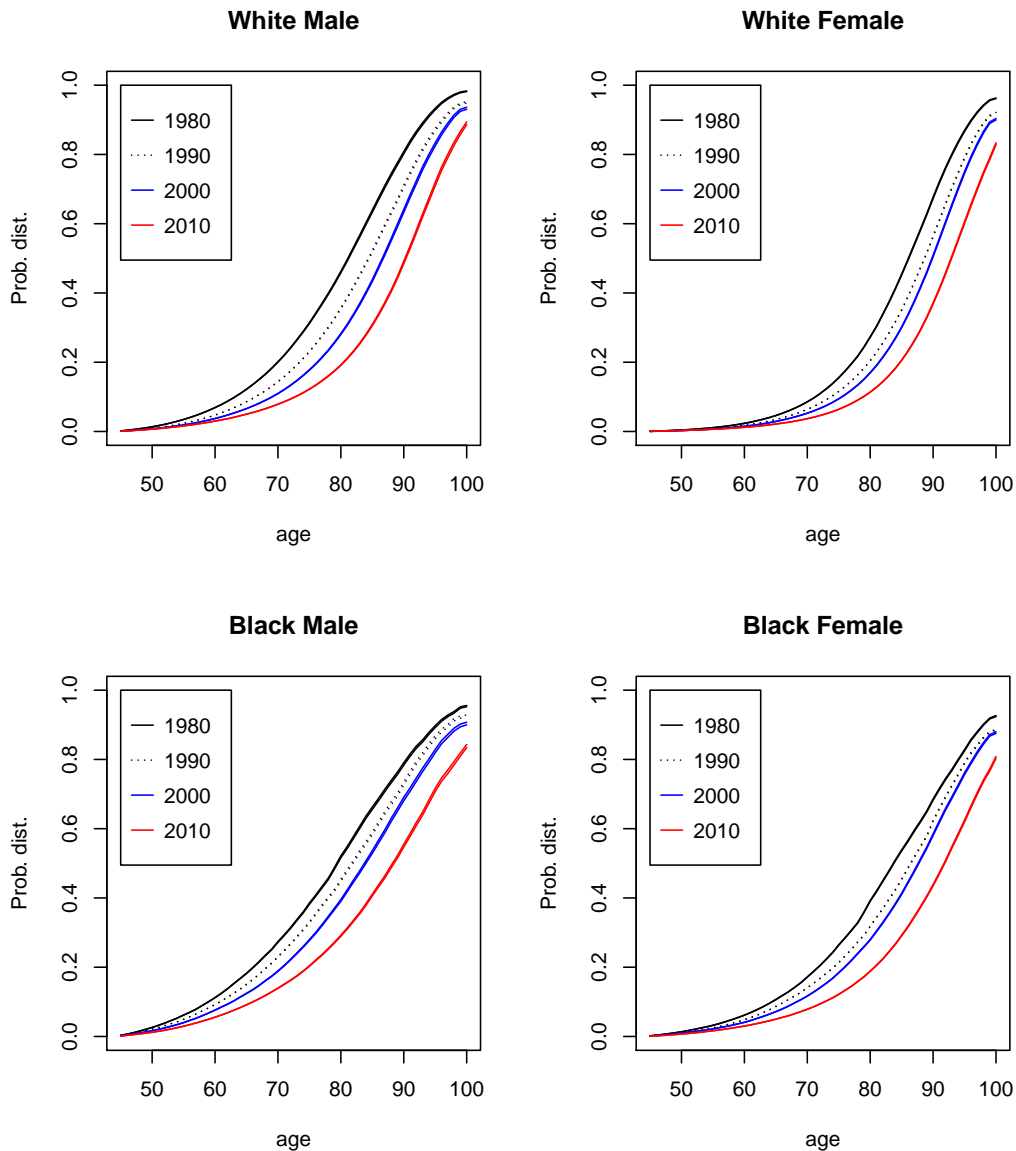
Koene *et al.* (2016), Blaes *et al.* (2017), Duarte *et al.* (2017) and references therein investigate common factors and possible association between cancer and CVD. Those studies suggest that genetic and behavioural factors such as smoking, drinking, diet, obesity, diabetes, and sitting times increase the risks of both diseases. Two risks are positively associated in the sense that they respond to shared factors in the same direction but with different magnitudes and patterns. For instant, long term exposure to air pollution increases the incidence of both diseases. CVD mortality responds to the level of exposure nonlinearly, whereas lung cancer mortality responds linearly (see Pope III *et al.*

Figure 3.3: Bounds on distribution functions of Cancer
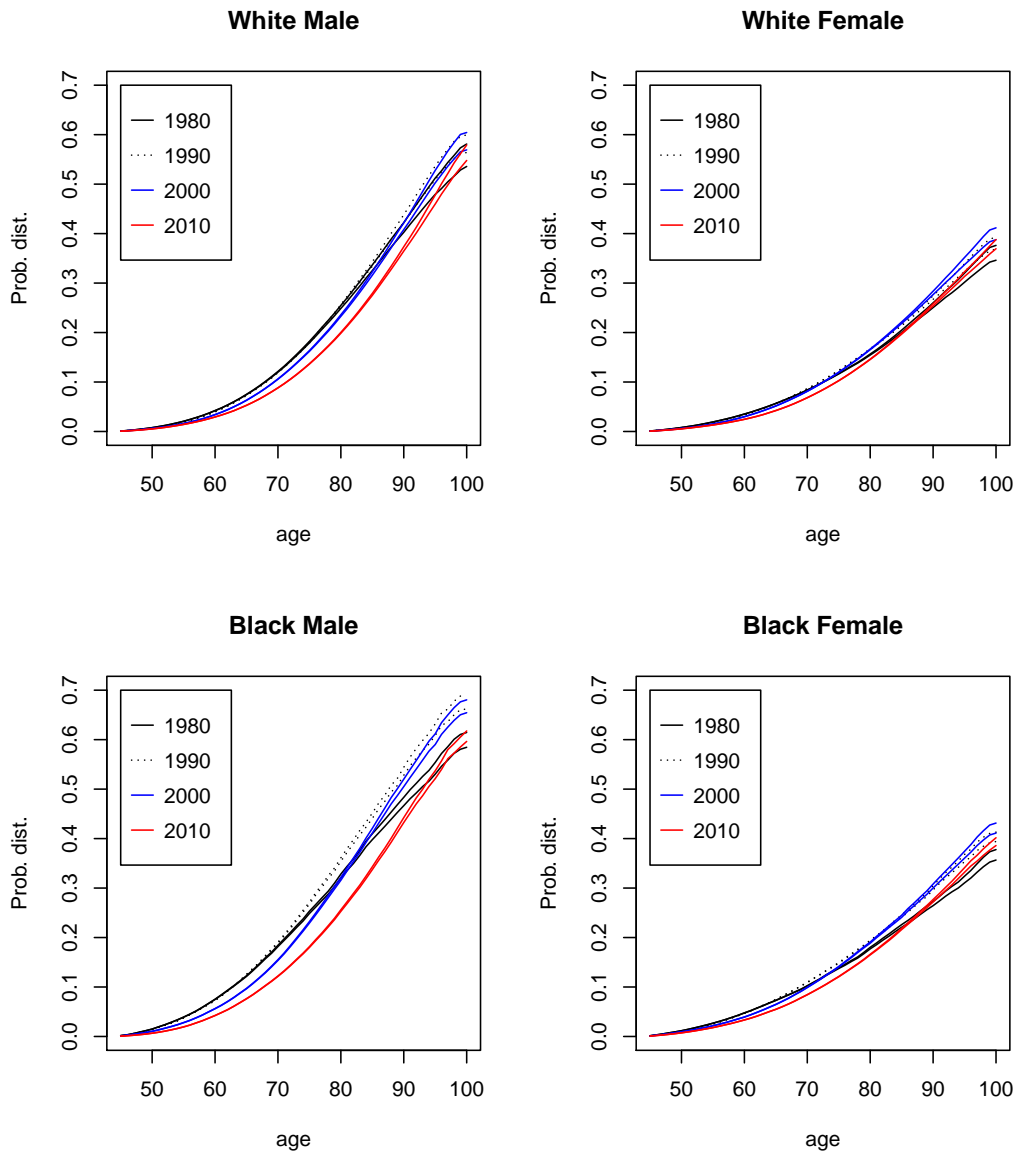
(2011)).[5] Much larger health effects are found for cancer than CVD. These results indicate that the two diseases are positively associated but the degree of association might not be strong.

Co-occurrence of those two diseases is not uncommon. Duarte *et al.* (2017) estimates that 20-30% of cancer patients also have a comorbid CVD. There is also evidence that cancer diagnosis (Fang *et al.* (2012)) and cancer therapies (Moslehi (2016)) increase the CVD risk. Regarding tail dependence, Driver

---

[5]CVD mortality sharply increases at the lower level of exposure but the slope becomes flatter as the level of exposure goes up. On the contrary, lung cancer mortality steeply and steadily increases over all levels of exposure.

**Table 3.3:** Bounds on the "progress ratio" under independence

group	1980-90	1980-2000	1980-2010
		CVD	
White Male	(1.085, 1.089)	(1.142, 1.147)	(1.233, 1.237)
White Female	(1.057, 1.059)	(1.087, 1.089)	(1.155, 1.157)
Black Male	(1.056, 1.063)	(1.110, 1.116)	(1.217, 1.223)
Black Female	(1.052, 1.054)	(1.082, 1.085)	(1.174, 1.176)
		Cancer	
White Male	(0.988, 0.998)	(1.001, 1.009)	(1.030, 1.036)
White Female	(0.990, 0.994)	(0.988, 0.993)	(1.005, 1.008)
Black Male	(0.961, 0.971)	(0.991, 1.000)	(1.047, 1.053)
Black Female	(0.982, 0.987)	(0.985, 0.989)	(1.005, 1.008)

et al. (2008) shows CVD and cancer mortality exhibit very different behaviour at advanced ages (over 80). This means that the Gumbel copula is less suitable for this application as it exhibits upper tail dependence.

I find different patterns of trends in mortality rates by allowing for positive association. To implement positive dependence between CVD and cancer, I use the Frank and Clayton copulas. Bounds are calculated with many different values of Kendall's $\tau \in \{0.01, 0.70\}$. Bounds for CVD are insensitive to the choice of the copula family as well as the degree of dependence so are omitted. On the contrary, bounds for cancer mortality are highly sensitive to the choice of the degree of dependence. This is due to the amount of dependent censoring. CVD is a more frequent hazard so it less suffers from censoring bias. Cancer is more frequently censored so that the bounds for cancer are more sensitive to the assumptions on dependence. Both bounds are not very sensitive to the choice of the copula family. The two different families of copulas provide very similar bounds given the same τ.

In general, larger reduction in cancer mortality is estimated under positive association than under independence. I first document bounds for cancer mortality in Figure 3.4 using the Frank copula with $\tau = 0.5$. For τ between 0 and 0.5, bounds look like in-between Figures 3.3 and 3.4. The mortality trends between 1980 and 2000 still exhibit heterogeneity across age groups and the younger have benefited more. Significant improvements are found in 2010 even at the right tail.

The progress ratios are displayed in Table 3.4. For men, these ratios are more than double the ratios under independence. The ratios for both white and black women are now around 5% which are much larger compared

Figure 3.4: Bounds on cancer mortality under Frank copula ($\tau = 0.5$)
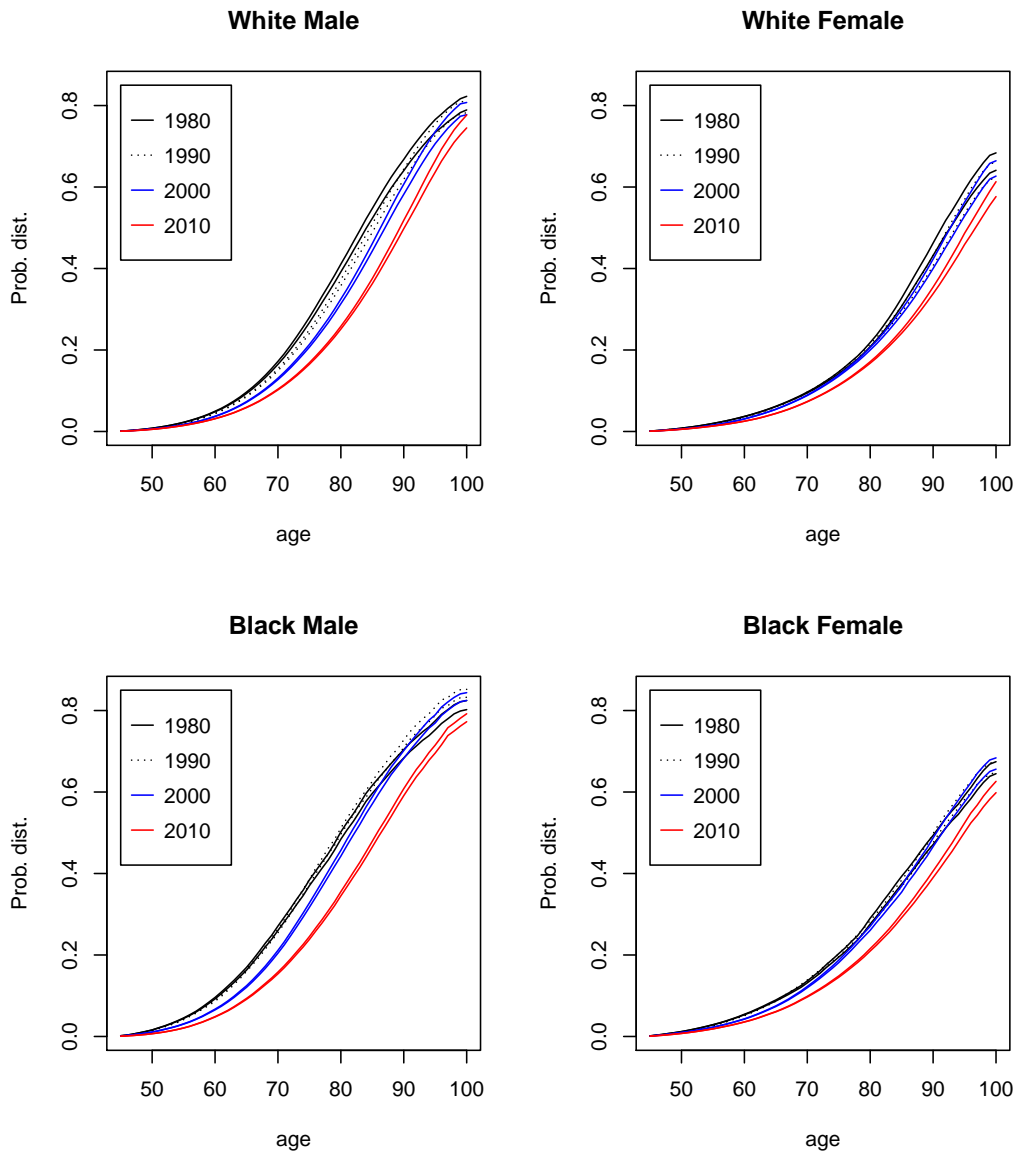
to below 1% under independence. Improvements in cancer mortality reach around a half of progress in CVD mortality. These results differ from HL's results. They estimated larger improvements for white and black females than their male counterparts. They also found white females have experienced the largest improvements. Bounds under both independence and dependence here show larger reduction for males than for females and so are more consistent with patterns in cancer statistics. The gender gaps shrink if dependence is allowed.

As mentioned, it is impossible to learn about τ from data. Given all

Table 3.4: Bounds on the "progress ratio" under Frank copula ($\tau = 0.5$)

| group	1980-90	1980-2000	1980-2010
	CVD		
White Male	(1.076, 1.083)	(1.135, 1.142)	(1.228, 1.236)
White Female	(1.051, 1.055)	(1.080, 1.084)	(1.151, 1.155)
Black Male	(1.044, 1.054)	(1.105, 1.115)	(1.222, 1.231)
Black Female	(1.045, 1.050)	(1.077, 1.082)	(1.172, 1.177)
	Cancer		
White Male	(1.014, 1.030)	(1.045, 1.060)	(1.100, 1.112)
White Female	(1.003, 1.016)	(1.010, 1.022)	(1.047, 1.056)
Black Male	(0.980, 0.996)	(1.026, 1.041)	(1.112, 1.124)
Black Female	(0.993, 1.005)	(1.004, 1.016)	(1.053, 1.062)

the results from different values of τ, the researcher must rely on a prior belief about the dependence. In this application, estimated improvements in cancer mortality are smallest under the independence assumption. Any degree of positive association delivers larger improvements. From the medical evidence, it is clear that those two hazards are positively associated. Therefore, estimated improvements under independence can be understood as the lower bound of improvements in cancer mortality. Likewise it is less likely that two risks are strongly correlated because both diseases respond to shared risk factors in different magnitudes and patterns. Hence the results in Table 3.4 can be regarded as the upper bound.

In Figure 3.5, I demonstrate that the progress ratios of cancer mortality for white men and women are quite monotone in τ. This is also the case for black men and women. The progress ratio for white males doubles the ratio relative to independence when τ is around 0.3. Near perfect association is required to attain HL's results for white females under dependence. However, there is no evidence that the dependence between two hazards are very different between males and females. The mortality rate from CVD is very flat across all values of τ so that the estimated progress in CVD mortality is robust to the copula choice.

Remark 5. *Differences between HL's results and mine are from methodologies, not from data. I try same exercises with the sample used by HL and main differences still preserve.*

Figure 3.5: Bounds on progress in expected survival times for white men and women
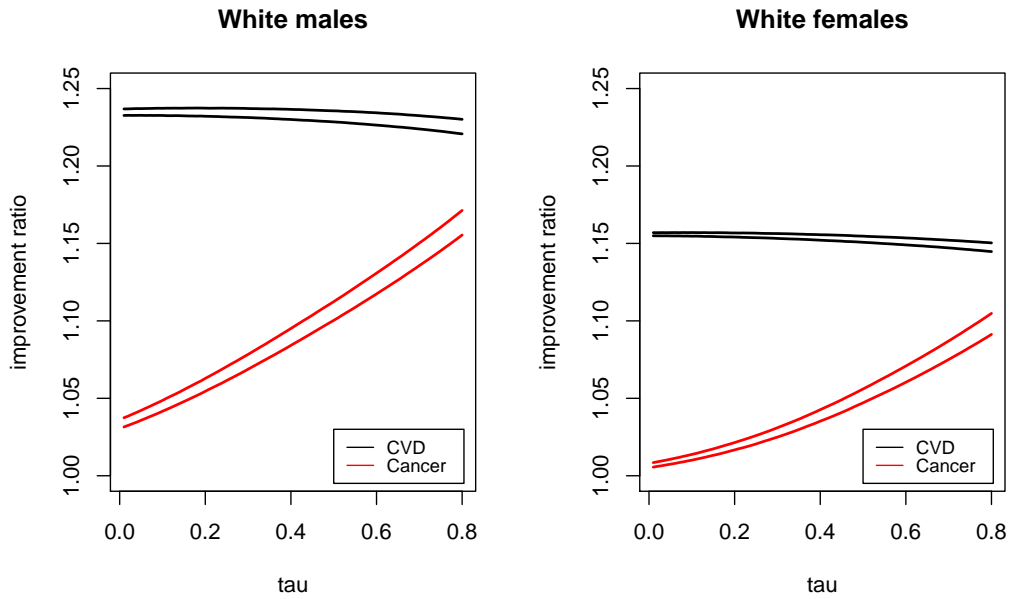
3.2.4   Counter-factual analysis

The bounds can be used to conduct counter-factual experiments. One important question is how reductions in cancer and CVD mortality contribute to improvements in the overall survival probability. This can be computed by fixing one of marginal distributions at the 1980 level and the other at the 2010 level. If one looks at the case where no reduction in cancer mortality has been achieved, using bounds in 1980 for cancer and bounds in 2010 for CVD provides the counter-factual survival probabilities given the copula. Then by comparing these probabilities to actual survival probabilities, the impact of reduction in CVD mortality on overall survival is understood.

Another interesting question is how much improvements in overall mortality would be achieved by further reduction in cancer or CVD mortality. Suppose the costs for marginal improvements in cancer and CVD mortality are the same. If a 20% reduction in cancer mortality leads to larger overall improvements than the same reduction in CVD mortality, we should invest more into cancer than CVD at the margin. Therefore, useful policy implications can be drawn from counter-factual experiments. Understanding overall survival rates is also important to operate pension schemes and national healthcare services.

I conduct counter-factual analysis with using the independent, Frank copulas. The latter copula is designed to have $\tau = 0.5$ so that it exhibit strong

Table 3.5: Counter-factual probability of surviving past 80 for the white

	White Male	White Female
Actual prob. (1980-2010)	0.404-0.647	0.614-0.757
	Independence copula	
No CVD reduction	[0.430, 0.434]	[0.620, 0.622]
No Cancer reduction	[0.602, 0.608]	[0.747, 0.750]
Further CVD reduction	[0.756, 0.758]	[0.833, 0.834]
Further Cancer reduction	[0.760, 0.762]	[0.844, 0.845]
	Frank copula : $\tau = 0.5$	
No CVD reduction	[0.433, 0.441]	[0.626, 0.631]
No Cancer reduction	[0.542, 0.559]	[0.723, 0.732]
Further CVD reduction	[0.726, 0.732]	[0.818, 0.822]
Further Cancer reduction	[0.725, 0.732]	[0.824, 0.828]

Table 3.6: Counter-factual probability of surviving past 80 for the black

	Black Male	Black Female
Actual prob. (1980-2010)	0.325-0.529	0.500-0.677
	Independence copula	
No CVD reduction	[0.356, 0.362]	[0.506, 0.509]
No Cancer reduction	[0.475, 0.483]	[0.665, 0.669]
Further CVD reduction	[0.669, 0.673]	[0.788, 0.789]
Further Cancer reduction	[0.648, 0.651]	[0.762, 0.764]
	Frank copula : $\tau = 0.5$	
No CVD reduction	[0.355, 0.363]	[0.514, 0.519]
No Cancer reduction	[0.437, 0.452]	[0.631, 0.642]
Further CVD reduction	[0.625, 0.634]	[0.764, 0.770]
Further Cancer reduction	[0.605, 0.613]	[0.737, 0.742]

positive association. Other copula families provide similar results given the same value of τ. I compute bounds on the counter-factual probabilities of surviving past 80 for each race-gender group. The results are shown in Table 3.5 for the white and Table 3.6 for the black. The top row displays the actual survival probability up to age 80 in 1980 and 2010. For each copula, the first and second rows report the counter-factual probabilities of survival until age 80 in the absence of reductions in CVD and cancer mortality since 1980 re-

spectively. The last two rows report the effects of 20% further reductions in CVD and cancer mortality from the 2010 level.

The overall survival probability until age 80 is improved mainly due to reduction in CVD mortality. Improvements in cancer mortality contributed little but its contribution increases when a strong positive relationship is assumed. In the last two rows, survival probabilities vary with the copula choice. Regardless of the choice, however, further reduction in CVD mortality leads to higher overall survival probabilities for black people. For white people, the benefits from improvements in both risks are almost identical.

3.2.5 Bounds for more than two hazards

So far I only use two hazards and but in the data four causes of death are available, CVD, lung cancer, the other cancers and all other causes. HL focused on only two hazards because including more risks is computationally very costly in their framework. But they pointed out that including more risks would be more desirable. Especially dividing cancer into its different types is important in the sense that lung cancer is the most common type of cancer and is mainly driven by smoking. Trends in lung cancer mortality are likely to exhibit different patterns from the other cancers. The proposed bounds in Chapter 2 are easy to compute even for a large number of risks.

I use three risks in the data, CVD, lung cancer, and the other cancers. Figure 3.6 shows the bounds for white males and females. Independence between durations is assumed. Bounds for CVD are identical to those of the bivariate risks model so are omitted. Bounds for the other cancers are very narrow but bounds for lung cancer become wider past age 90 since lung cancer is the smallest risk. White males experience reductions in both types of cancers. For white females, reduction in the other cancers is comparable to that of white males. On the contrary, lung cancer mortality deteriorated over time and stayed at the worst level until 2010. Similar patterns are observed for black men and women. The progress ratios in Table 3.7 show the same patterns. Around 2-3% improvements in the other cancers are quite similar across all demographic groups. White and black women experience increases in lung cancer mortality. For both white and black men, reductions in cancer mortality do not differ between two types of cancer. These results are sensible considering that female smoking peaked in 1970s, whereas male smoking peaked around 10 years earlier (Lomborg (2003)).

Allowing for dependence among risks amplifies progress in cancer mortality and the gender gap. Table 3.8 shows the progress ratios computed from the

Figure 3.6: Bounds on distribution functions for white men and women
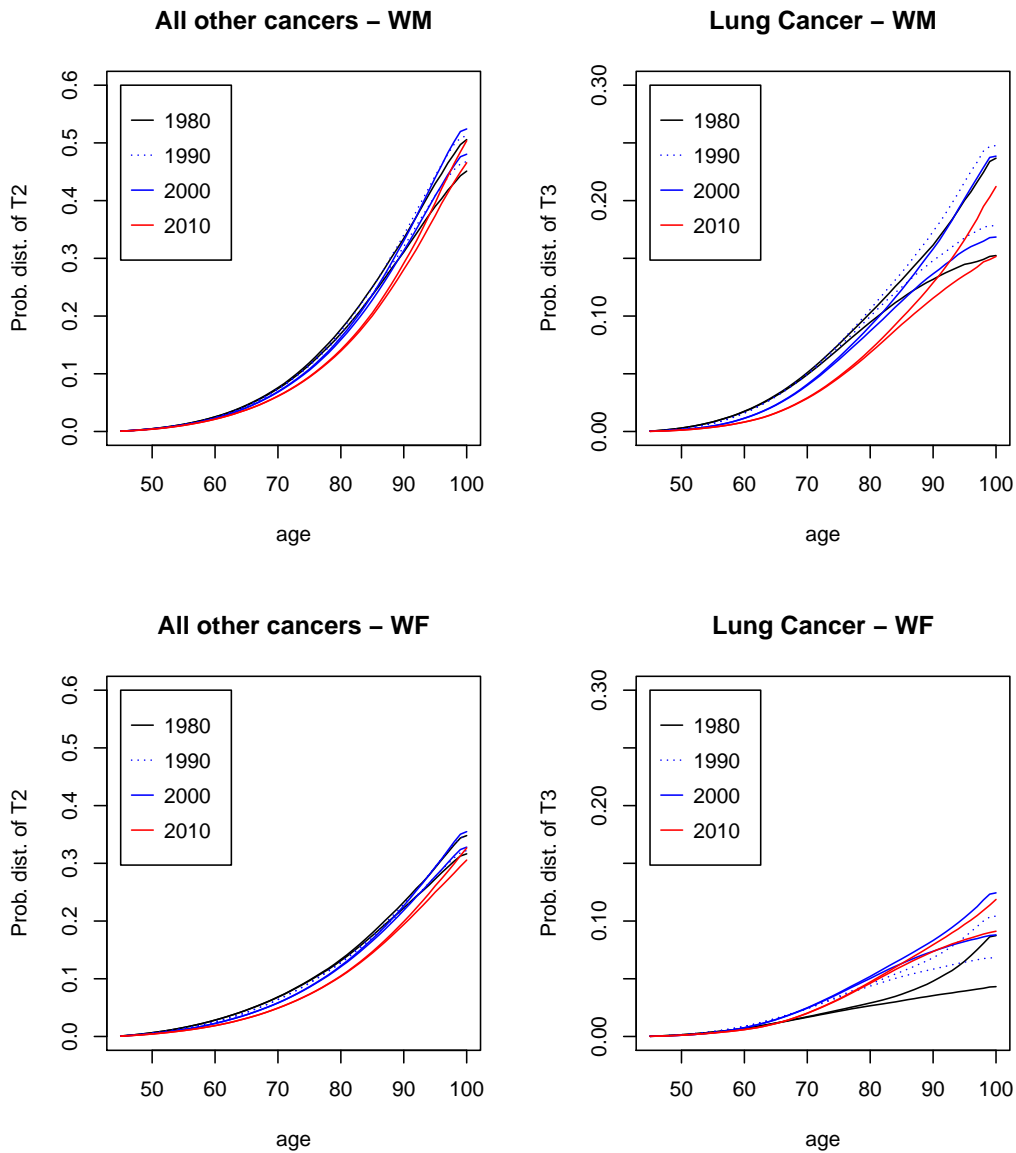
Frank copula with $\tau = 0.5$. For both white and black men, progress ratios for all the other cancer are around 10% but females only have less than two third of it. Progress ratios in lung cancer mortality also go up. Bounds are quite wide but it is clear that improvements in lung cancer mortality for females are very small even when quite strong association between risks is assumed. These results partly explain why females experience much less progress in overall cancer mortality in the two risks case. A large fraction of the gender gap in cancer mortality is due to the gap in lung cancer mortality.

Table 3.7: Bounds on the "progress ratio" under the independent copula

group	1980-90	1980-2000	1980-2010
All other cancers			
White Male	(0.992, 1.003)	(0.997, 1.007)	(1.016, 1.023)
White Female	(1.000, 1.005)	(1.004, 1.008)	(1.019, 1.022)
Black Male	(0.970, 0.982)	(0.987, 0.998)	(1.028, 1.035)
Black Female	(0.991, 0.996)	(0.997, 1.002)	(1.017, 1.020)
Lung cancer			
White Male	(0.988, 1.001)	(0.998, 1.010)	(1.014, 1.023)
White Female	(0.986, 0.992)	(0.981, 0.986)	(0.984, 0.988)
Black Male	(0.974, 0.990)	(0.991, 1.005)	(1.017, 1.026)
Black Female	(0.987, 0.992)	(0.983, 0.988)	(0.986, 0.989)

Table 3.8: Bounds on the "progress ratio" under Frank copula ($\tau = 0.5$)

group	1980-90	1980-2000	1980-2010
All other cancers			
White Male	(1.015, 1.033)	(1.043, 1.059)	(1.094, 1.108)
White Female	(1.010, 1.022)	(1.020, 1.031)	(1.057, 1.067)
Black Male	(0.983, 1.002)	(1.024, 1.041)	(1.103, 1.117)
Black Female	(0.999, 1.010)	(1.012, 1.023)	(1.061, 1.071)
Lung cancer			
White Male	(0.993, 1.038)	(1.022, 1.066)	(1.073, 1.111)
White Female	(0.976, 1.020)	(0.975, 1.012)	(1.003, 1.031)
Black Male	(0.962, 1.011)	(0.999, 1.044)	(1.073, 1.111)
Black Female	(0.967, 1.013)	(0.968, 1.008)	(1.002, 1.030)

3.3   Unemployment spells

There is some agreement that higher unemployment benefits lead to longer unemployment spells (see Card *et al.* (2015)), and estimating the magnitudes of the responses is of concern to policy-makers. In this application, I focus on the exercises in Farber *et al.* (2015) (FRV henceforth) who study effects of extended unemployment insurance (UI) benefits on unemployment spells during the Great Recession and its aftermath in the US. They also estimate the impact of extended UI benefits on probabilities of re-employment and exit from the labour force. The main result of their paper is that extension of UI benefits did not affect exit through re-employment, whereas it discouraged exit from the labour force.

This exercise provides important policy implications. A long unemployment duration itself has negative effects on an unemployed person's chance of

job finding via multiple channels as shown in Kroft *et al.* (2013) and references therein. Therefore, if extended UI benefits indeed prolonged unemployment spells of active job seekers by deter their motivations, this policy would cause adverse effects on efficiency of an economy. On the other hand, if it only attracted unemployed people who were more likely to exit from the labour force, extended benefits would not bring inefficiency.

Farber *et al.* (2015) and FRV only focus on exit rates from unemployment using binary choice (logit) models. They do not investigate the effects of extended benefits on unemployment durations. Furthermore, they treat the other type of exit as independent censoring when they use binary response models for each type of exit from unemployment. However, durations up to re-employment and exit from the labour force may be dependent. Suppose an unemployed individual is highly motivated to find a job. Her job search effort is likely intense and her unemployment spell up to re-employment is likely short. On the other hand, her counter-factual duration up to exit from the labour force would be long. For an individual with no strong motivation, it would be the other way round. The independence assumption may not deliver a correct answer in such a sense.

I investigate effects of extended UI benefits on distributions of unemployed spells up to two types of exits using data from FRV. Sensitivity analysis on the independence assumption is conducted by allowing for negative dependence between two durations. The data is a sample of 56,491 unemployment spells from Current Population Survey (CPS) for 2008-2014. All individuals in the sample were aged 18-69 and were potentially eligible for UI benefits. The data do not include information about actual receipt of UI benefits. Therefore, I maintain the same assumptions in FRV that every unemployed person in the sample is eligible for benefits and receives UI benefits from the date of displacement to the maximum duration.

Most states in the US extended UI benefits up to 99 weeks from the usual duration (26 weeks) between 2008-2010. These extended durations were tapered from the first quarter of 2012 so that no state provided UI benefits beyond the usual duration in 2014. To investigate effects of extended benefits, I divide the sample into two periods, 2008-2011 (phase-in) and 2012-2014 (phase-out). Those periods reflect expansion and contraction of the benefit duration as well as increases and decreases of labour market slackness. Therefore, the distributions of unemployment durations in two periods reflect the benefit durations and the labour market conditions. However, economic theories and empirical evidence in the literature suggest that extended UI benefits have

negative effects on the job finding rate and the probability of exit from the labour force. More favourable labour market conditions bolster the job finding rate but deter an unemployed person from exiting the labour force. These presumptions can be used to derive insights on separate effects of labour market slack and UI benefits on both durations.

Figure 3.7: Empirical distribution of unemployment durations
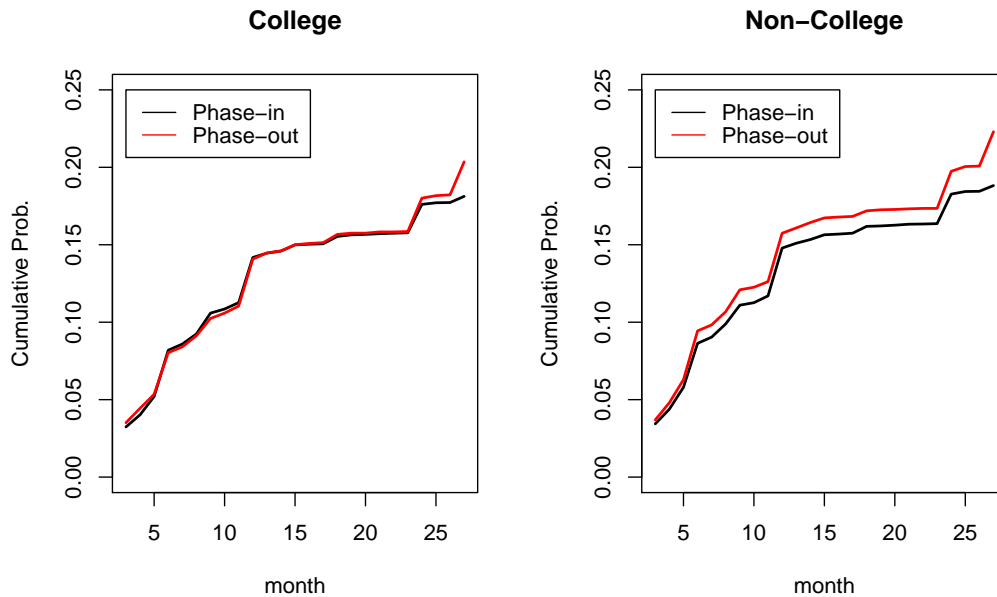
The empirical distribution of unemployment durations in each period shows that unemployed individuals exit from unemployment more quickly in the phase-out period. Figure 3.7 displays the empirical distributions for college graduates (including postgraduates) and non-college graduates. The two groups exhibit somewhat different patterns. The distribution of college graduates does not vary between periods much. Only the long term unemployed (2 years or longer) were more likely to exit in the phase-out period. On the contrary, non-college graduates were more likely to leave unemployment in the phase-out period with a significant margin and this margin constantly increases as the unemployment duration becomes longer. Workers with higher education levels tend to face a lower unemployment rate and shorter unemployment spells as empirically shown in Nickell (1979). The same pattern is observed in the recent work (see Kroft *et al.* (2016)).

The separate effects of UI benefits on exit from the labour force and re-employment are of interest. I use independent and Frank copulas with negative values of τ to derive bounds on the distributions of latent spells. Gender is also

Figure 3.8: Bounds on distributions for males in phase-in and phase-out periods
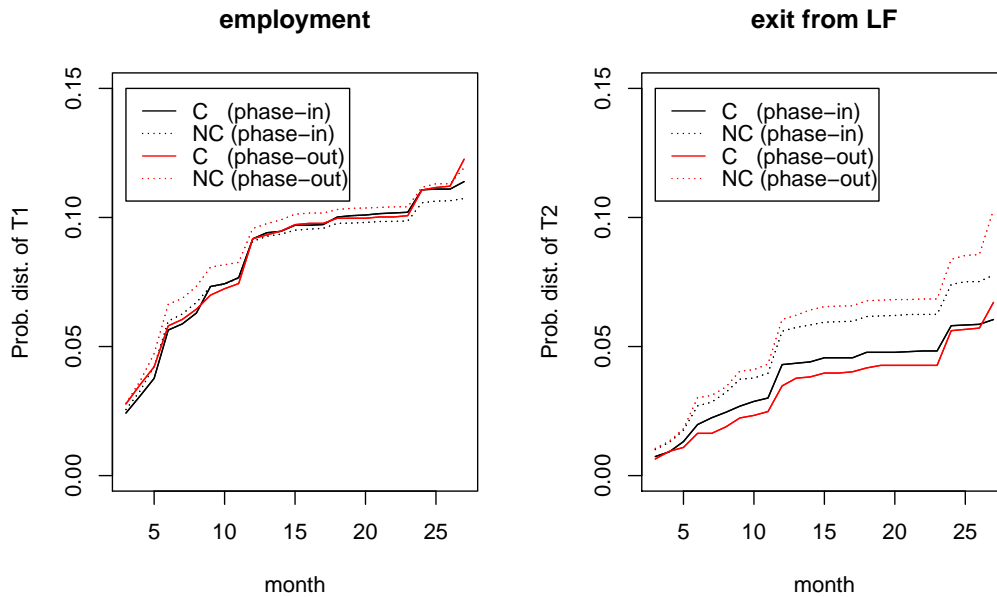
considered as a covariate. Allowing for negative dependence between durations does not alter the bounds much. Bounds are slightly downward shifted when negative dependence is allowed. All the following bounds are computed from the Frank copula with $\tau = -0.5$. Those bounds are shown to be very tight.

Figure 3.8 shows the bounds for male college graduates and non-college graduates in the two periods. College graduates' distribution of the duration up to re-employment is unchanged over the two periods but the distribution of duration until exit from the labour force was downward shifted in the phase-out period except for the long-term unemployed. Male non-college graduates exhibit different patterns. Their distributions of durations up to re-employment and exit from the labour force were upward shifted. Given the more favourable labour market situation in the phase-out period, the effects of extended benefits were marginal on durations up to re-employment for both college graduates and non-college graduates. For exit from the labour force, the negative effects of tapering of extended benefits dominated the positive effects of better labour market conditions for non-college graduates and vice versa for college graduates.

For females, the patterns are more obvious as displayed in Figure 3.9. Durations up to re-employment were affected very little, whereas the distributions of duration until exit from the labour force were all upward shifted with significant margins which grow as the unemployment duration becomes longer.
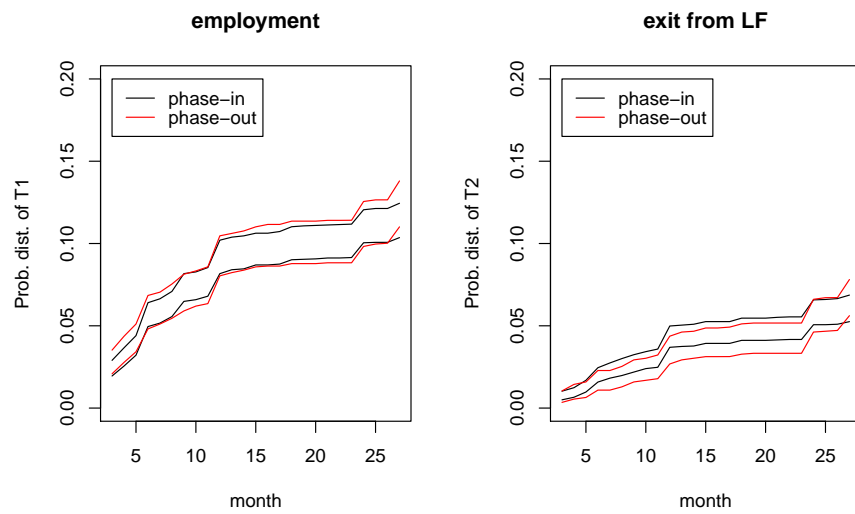
Figure 3.9: Bounds on distributions for females in phase-in and phase-out periods
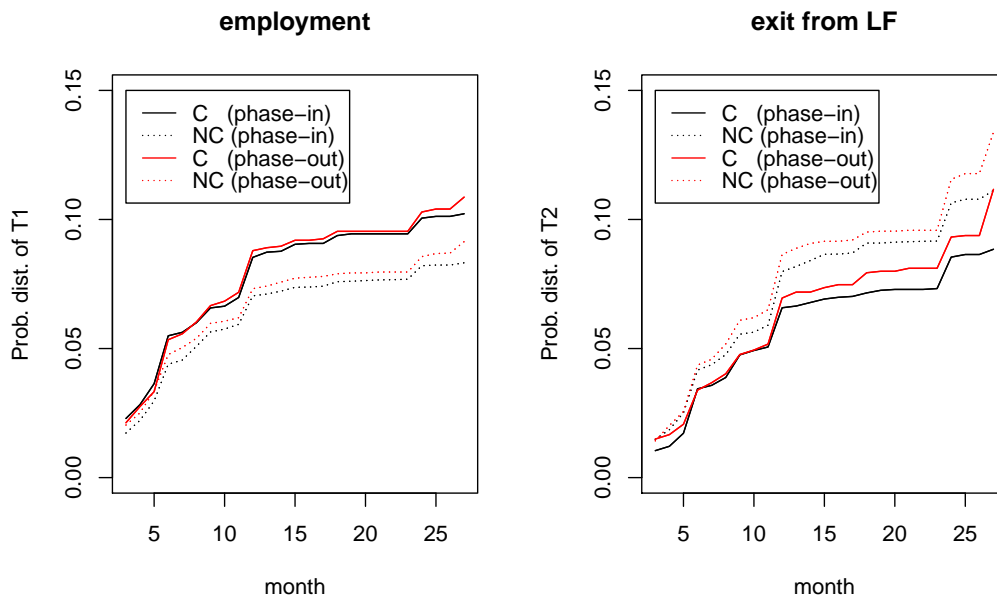
These results show that the negative effects of roll back of extended benefits were significantly large so that the effects of tighter labour market conditions were dominated for unemployed females. The different patterns between men and women may reflect the gender difference in the job industry distributions. Women were more concentrated in the service sector but men were much more dispersed over many sectors.

The bootstrap inference method is employed to test whether the estimated patterns are statistically significant. Confidence bands for bound estimates are displayed in Figure 3.10-3.13. At 95% confidence level, the pattern that non-college graduates with long (≥ 2 years) unemployment durations were more likely to exit from the labour force in the phase-out period is significant for both men and women. This pattern remains significant at the same confidence level when educational level is uncontrolled.

These exercises show that UI benefits affected the duration until exit from the labour force but not the duration up to re-employment. Negative effects of UI benefits were larger for non-college graduates and females. Whether the extended UI benefits discourage active job seekers has been a long lasting question. The main results here support the original findings in FRV that UI benefits did not play a pivotal role to fuel moral hazard of the unemployed people. The extension of benefits suppressed exit from the labour force during the phase-in period but it did not distort the efficiency of the US economy as

Figure 3.10: 95% confidence bands on distributions for male college graduates in phase-in and phase-out periods

Figure 3.11: 95% confidence bands on distributions for male non-college graduates in phase-in and phase-out periods
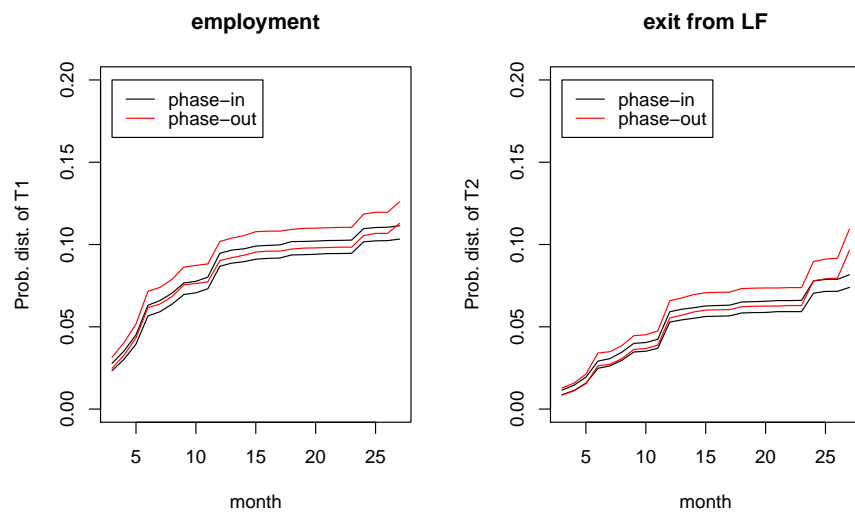
it did not deter active job seekers from finding a job.

3.4 Conclusion

I employ the bounds approach in empirical applications to the war on cancer and to unemployment spells. These applications illustrate the usefulness of the proposed approach. It is easy to implement and is widely applicable in many applications in economics and other applied studies. As the approach

Figure 3.12: 95% confidence bands on distributions for female college graduates in phase-in and phase-out periods
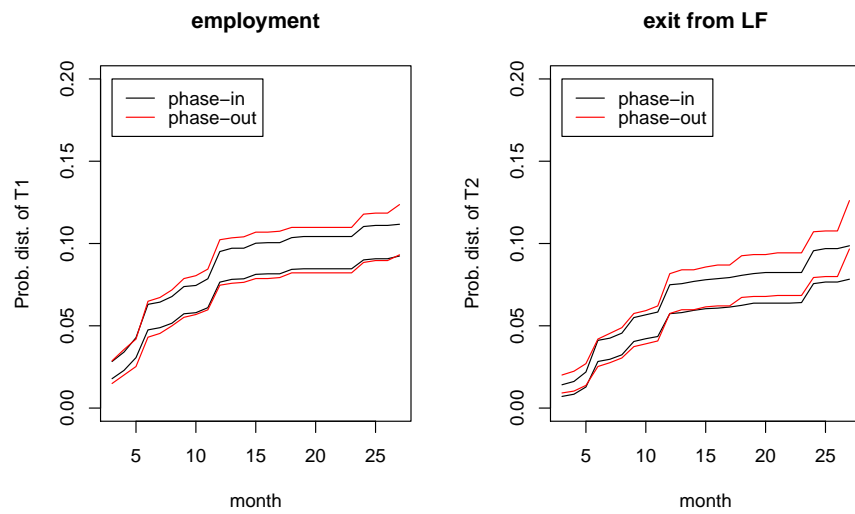
Figure 3.13: 95% confidence bands on distributions for female non-college graduates in phase-in and phase-out periods
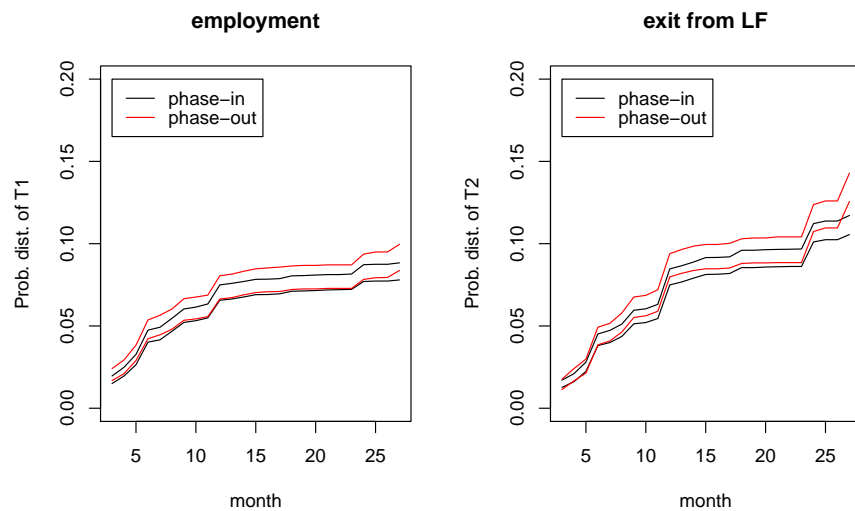
is computationally very attractive, a bootstrap inference procedure provides confidence bands for bounds on distribution functions without requiring much computational burden.

I re-evaluate trends in CVD and cancer mortality. By allowing for dependence between the two risks, I find that reductions in cancer mortality are larger than previously shown. Estimated patterns differ from the findings in Honoré and Lleras-Muney (2006). The model is also extended to include more

than 2 hazards. Dividing cancers into subcategories reveals discover different trends in mortality rates from lung cancer and the other cancers. In another application, I study the effects of extended UI benefits on the distributions of unemployment durations. The estimated bounds support the original findings in Farber *et al.* (2015) that extended UI benefits did not distort efficiency of the US economy much.

The application to cancer mortality offers interesting extensions for future research. One important extension would be to analyse the relationship between mortality trends and regional characteristics such as wealth, income and other demographic factors. The US Census and Multiple Cause of Death data provide state/county level population information and death counts. By conducting the mortality analysis at a county/state level using the bounds approach, interesting empirical findings can be revealed.

Appendix A

Supplemental materials for Chapter 1

Elucidation of computational details

Under parametric restrictions, the algorithm introduced in Section 1.4 provides the identified set. Computational details of the identified set are provided in this section. Given a parameter grid θ_j, the threshold functions $p_y(x)$ for all x and y are generated. Therefore, the ordering between threshold values are θ_j specific. Given the thresholds, define that for all $y \in \mathcal{R}_Y \backslash \{0\}$ and for all $x, x' \in \mathcal{R}_X$,

$$\phi_1(y, x', x) \equiv \max\{m : p_m(x') \leq p_y(x)\}, \quad \phi_2(y, x', x) \equiv \min\{m : p_y(x) \leq p_m(x')\}.$$

Since $U \sim Unif(0, 1)$, for any interval $\mathcal{S}$, $G_U(\mathcal{S})$ is equal to the length of $\mathcal{S}$. This means that $G_U([0, p_y(x)]) = p_y(x)$ and $G_U([p_y(x), 1]) = 1 - p_y(x)$. The containment functional given the interval $[0, p_y(x)]$ is

$$C_h([0, p_y(x)]|z) = P[\mathcal{U}(Y, X; h) \subseteq [0, p_y(x)] \mid z]. \tag{A.1}$$

When $X = x$, intervals $[0, p_1(x)], \cdots, [p_{y-1}(x), p_y(x)]$ are contained in $[0, p_y(x)]$. For $X = x'$ where $x' \neq x$, intervals $[0, p_1(x')], \cdots, [p_{\phi_1(y,x',x)-1}(x), p_{\phi_1(y,x',x)}(x)]$ are contained in $[0, p_y(x)]$. Therefore, noticing that $\phi_1(y, x', x) = y$, the containment functional is

$$C_h([0, p_y(x)]|z) = \sum_{x' \in \mathcal{R}_X} P[Y \leq \phi_1(y, x', x) - 1 \ \wedge \ X = x' \mid z]. \tag{A.2}$$

By similar logic, noticing that $\phi_2(y, x, x) = y$, the containment functional given

$[p_y(x), 1]$ is

$$C_h([p_y(x), 1]|z) = \sum_{x' \in \mathcal{R}_X} P[Y \geq \phi_2(y, x', x) \ \wedge \ X = x' \mid z]. \qquad \text{(A.3)}$$

Now by the inequality (1.11), the upper and lower bounds in Step 4 of the algorithm are respectively,

$$LB(y, x, z) \equiv \sum_{x' \in \mathcal{R}_X} P[Y \leq \phi_1(y, x', x) - 1 \ \wedge \ X = x' \mid z] \leq p_y(x) \qquad \text{(A.4)}$$

$$UB(y, x, z) \equiv \sum_{x' \in \mathcal{R}_X} P[Y \leq \phi_2(y, x', x) - 1 \ \wedge \ X = x' \mid z] \geq p_y(x) \qquad \text{(A.5)}$$

and these inequalities hold for all values of $z \in \mathcal{R}_Z$.

Additional heterogeneity

Unobserved heterogeneity U is so far assumed to be a random scalar. Suppose some elements of X are unobserved. Those elements are relevant in the structural function of Y but omitted. This may happen in practice so it needs to be accommodated in the model (1.9). Let U be a vector and suppose that U is two dimensional unobserved heterogeneity such that $U := (U_1, U_2)$ where $U_1 \sim Unif(0, 1)$ and U_2 is continuously distributed with its density function $f_{U_2}(\cdot)$. Then the model (1.9) is modified as follows.

$$
\begin{aligned}
Y = h(X, U) = \quad & 0 \quad && \text{if } p_0(X, U_2) && \leq U_1 \leq && p_1(X, U_2) \\
= \quad & 1 \quad && \text{if } p_1(X, U_2) && < U_1 \leq && p_2(X, U_2) \\
= \quad & \cdots && && && \\
= \quad & y \quad && \text{if } p_y(X, U_2) && < U_1 \leq && p_{y+1}(X, U_2) \\
= \quad & \cdots && && &&
\end{aligned}
\qquad \text{(A.6)}
$$

where $p_0(X, U_2)$ is normalized to 0. If U_2 is observable, then the set of threshold functions $\{p_{y+1}(x, u_2)\}_{y \in \mathcal{R}_Y, \ u_2 \in \mathcal{R}_{U_2}, \ x \in \mathcal{R}_X}$ is the object of identification.

Suppose $X \perp\!\!\!\perp U$ and $U_1 \perp\!\!\!\perp U_2$. Without observing U_2, there is no hope of identifying $p_{y+1}(x, u_2)$. The threshold functions are naturally specified by

$$p_{y+1}(x, u_2) = \mathbb{P}[Y \leq y | X = x, U_2 = u_2].$$

This probability cannot be identified but the average threshold functions

$$p_{y+1}(x) \equiv \int_{\mathcal{R}_{u_2}} p_{y+1}(x, u_2) f_{U_2}(u_2) du_2$$

are point identified by observing $\mathbb{P}[Y \leq y | X = x]$.

In the case where X is not independent of U, $p_{y+1}(x)$ is not point identified because now

$$\mathbb{P}[Y \leq y | X = x] = \int_{\mathcal{R}_{u_2}} p_{y+1}(x, u_2) f_{U_2|X}(u_2|x) du_2.$$

It can be partially identified under the existence of an instrument Z which satisfies $Z \perp\!\!\!\perp U$.

Given the values of Y and X, the level set $\mathcal{U}(y, x; h)$ is derived on $\mathcal{R}_U \equiv [0, 1] \times \mathcal{R}_{u_2}$ such that

$$\mathcal{U}(y, x; h) = \{([p_y(x, u_2), p_{y+1}(x, u_2)], u_2) : u_2 \in \mathcal{R}_{u_2}\}.$$

Let $\mathsf{U}_h(x)$ denote the conditional support of the $\mathcal{U}$-level set given $X = x$. Then the unconditional support of the $\mathcal{U}$-level set is $\mathsf{U}_h \equiv \{\mathsf{U}_h(x) : x \in \mathcal{R}_X\}$. Let $\tilde{\mathsf{Q}}_h$ be the collection of all the connected unions of elements of U_h. Then $\tilde{\mathsf{Q}}_h$ is the collection of CDTS.

Corollary 3. *Suppose $Z \perp\!\!\!\perp U$. Given the joint distribution of (Y, X, Z) and the model (A.6), the identified set for the structural function h is characterized as follows.*

$$\mathcal{H}^* = \{h : \forall \mathcal{S} \in \tilde{\mathsf{Q}}_h, \quad C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \ a.e \ z \in \mathcal{R}_Z\}.$$

Assume that $U_1 \perp\!\!\!\perp U_2$. $G_U(\mathcal{S})$ can be computed given $\mathcal{S}$. For example, suppose $\mathcal{S} = \{([p_y(x, u_2), p_{y+1}(x, u_2)], u_2) : u_2 \in \mathcal{R}_{u_2}\}$. Then

$$G_U(\mathcal{S}) = \int_{\mathcal{R}_{U_2}} [p_{y+1}(x, u_2) - p_y(x, u_2)] f_{U_2}(u_2) du_2.$$

Likewise, the containment functional is

$$C_h(\mathcal{S}|z) = P[\mathcal{U} \in \mathcal{S}|z].$$

Therefore, Corollary 3 implies

$$P[\mathcal{U} \in \mathcal{S}|z] \leq p_{y+1}(x) - p_y(x).$$

Repeating this procedure for all elements in $\tilde{\mathsf{Q}}_h$ yields bounds for the average threshold functions.

Proofs of main results

All the corollaries are direct applications of the results in the paper so the proofs are omitted.

Proof of Proposition 1. Suppose that $\mathcal{R}_X = \{0, 1, 2, \cdots, K\}$ without loss of generality. If there exist $x, x' \in \mathcal{R}_X$ and $y, y' \in \mathcal{R}_Y$ such that $y - \exp(\alpha + \beta x) = y' - \exp(\alpha + \beta x') = \bar{u}$, then $f_U(\bar{u}) = \mathbb{P}[Y = y \cap X = x|Z] + \mathbb{P}[Y = y' \cap X = x'|Z]$ and the probability distribution of X can vary with Z. Suppose that $x > x'$ and define $h = y - y'$ then $\alpha = \ln \frac{h}{\exp(\beta x) - \exp(\beta x')}$. Given x, x', h and β, the value of α satisfying the above equation is found unless the signs of h and $\exp(\beta x) - \exp(\beta x')$ are different. Therefore, a set of pairs (α, β) given (h, x, x') is found, which is a curve on $\mathbb{R}^2$. As X and Y are discrete, the number of curves is countably infinitely many as so are the number of possible combinations of (h, x, x'). $\hspace{1cm}$ Q.E.D.

Proof of Proposition 2. Suppose that $\mathcal{R}_X = \{0, 1, 2, \cdots, K\}$. If there exist $x, x' \in \mathcal{R}_X$ and $y, y' \in \mathcal{R}_Y$ such that $y/\exp(\alpha + \beta x) = y'/\exp(\alpha + \beta x') = \bar{u}$, then $f_U(\bar{u}) = \mathbb{P}[Y = y \cap X = x|Z] + \mathbb{P}[Y = y' \cap X = x'|Z]$ and the probability distribution of X can vary with Z. Define $h = \frac{y}{y'}$ where $y' \neq 0$, then

$$h = \frac{\exp(\alpha + \beta x)}{\exp(\alpha + \beta x')} = \exp(\beta(x - x')) \implies \beta = \frac{\log h}{x - x'} \qquad \text{(A.1)}$$

Given x, x' and h, a unique value of β satisfies the above equation. Therefore, a set of pairs (α, β) satisfying (A.1) is a horizontal line on $\mathbb{R}^2$. As X and Y are discrete, the number of curves is countably infinitely many as so are the number of possible combinations of (h, x, x') $\hspace{1cm}$ Q.E.D.

Proof of Theorem 1. Define that for all $y \in \mathcal{R}_Y \backslash \{0\}$ and for all $x, x' \in \mathcal{R}_X$,

$$\rho_{xy}(z) \equiv \mathbb{P}[X = x \ \cap \ Y = y \mid z].$$

$\phi_1(y, x', x)$ and $\phi_2(y, x', x)$ are defined in Appendix I. Since $U \sim Unif(0, 1)$,

$$G_U([0, p_y(x)]) = p_y(x), \quad G_U([p_y(x), 1]) = 1 - p_y(x).$$

The containment functionals given the intervals are

$$C_h([0, p_y(x)]|z) = \sum_{x' \in \mathcal{R}_X} \sum_{m=0}^{\phi_1(y,x',x)-1} \rho_{x'm}(z) \qquad (A.2)$$

$$C_h([p_y(x), 1]|z) = \sum_{x' \in \mathcal{R}_X} \sum_{m=\phi_2(y,x',x)}^{\infty} \rho_{x'm}(z) \qquad (A.3)$$

Given the inequality (1.11), the upper and lower bounds of $p_y(x)$ given z are found. Then by the intersection of these bounds across z,

$$\sup_{z \in \mathcal{R}_Z} C_h([0, p_y(x)]|z) \le p_y(x) \le \inf_{z \in \mathcal{R}_Z} \{1 - C_h([p_y(x), 1]|z)\}. \qquad (A.4)$$

Q_h is core determining if no additional interval in $\tilde{\mathsf{Q}}_h$ makes the above bounds tighter. By any additional interval $[p_y(x), p_k(x')]$ where $p_{y+1}(x) \le p_k(x')$ for $y, k \in \mathcal{R}_Y \backslash \{0\}$ and $x, x' \in \mathcal{R}_X$, the lower bound of $p_k(x') - p_y(x)$ is delivered as follows.

$$\sup_{z \in \mathcal{R}_Z} \left\{ \sum_{x'' \in \mathcal{R}_X} \sum_{m=\phi_2(y,x'',x)}^{\phi_1(k,x'',x')-1} \rho_{x''m}(z) \right\} =$$
$$\sup_{z \in \mathcal{R}_Z} \left\{ \sum_{x'' \in \mathcal{R}_X} \sum_{m=0}^{\phi_1(k,x'',x')-1} \rho_{x''m}(z) - \sum_{x'' \ne x} \sum_{m=0}^{\phi_2(y,x'',x)-1} \rho_{x''m}(z) \right\} \qquad (A.5)$$

Under Condition 1, for all $i = \{1, 2, \cdots, K\}$ and y, $[p_y(x_i), p_{y+1}(x_i)]$ contains $\{p_y(x_j) : i < j, \ j \in \{1, \cdots, K\}\}$ and $\{p_{y+1}(x_j) : i > j, \ j \in \{1, \cdots, K\}\}$. For all intervals $[p_y(x), p_{y+1}(x)] \in \mathsf{U}_h$, there exists $s \in \mathcal{R}_Y$ such that $p_s(x') \in [p_y(x), p_{y+1}(x)]$ for all x' and $\phi_1(k, x'', x') \ge \phi_2(y, x'', x)$ is guaranteed for all x''. Therefore, the equality in (A.5) holds.

From the inequalities attained by Q_h, the lower bound of $p_k(x') - p_y(x)$ is also constructed. By subtracting the upper bound of $p_y(x)$ from the lower bound of $p_k(x')$, the lower bound of $p_k(x') - p_y(x)$ is

$$\sup_{z \in \mathcal{R}_Z} \left\{ \sum_{x'' \in \mathcal{R}_X} \sum_{m=0}^{\phi_1(k,x'',x')-1} \rho_{x''m}(z) \right\} - \inf_{z \in \mathcal{R}_Z} \left\{ \sum_{x'' \ne x} \sum_{m=0}^{\phi_2(y,x'',x)-1} \rho_{x''m}(z) \right\}$$

This bound is weakly tighter than (A.5) and thus it is redundant to check $[p_y(x), p_k(x')] \in \tilde{\mathsf{Q}}_h/\mathsf{Q}_h$. Therefore, Q_h is core determining.   Q.E.D.

Appendix B

Supplemental materials for Chapter 2

Bounds for many hazards

Let $p_{m,j}^U(x)$ and $p_{m,j}^L(x)$ denote upper and lower bounds of $p_{m,j}(x)$. Given that C is continuous and increasing in its arguments, the volume function $V_C([\mathbf{a}, \mathbf{1}])$ is strictly monotone in the j-th element of $\mathbf{a}$. Therefore, one can define the inverse of V_C for a_j. Suppose that $V_C([\mathbf{a}, \mathbf{1}]) = c$ for some constant $c \in [0, 1]$. Then there exists a function $V_{Cj}^{-1} : [0, 1]^J \to [0, 1]$ such that

$$V_{Cj}^{-1}(\mathbf{a}_{-j}, c) = a_j \quad \text{where} \quad \mathbf{a}_{-j} = (a_1, \cdots, a_{j-1}, a_{j+1}, \cdots, a_J).$$

Define

$$\mathbf{a}_{m,j}^L(x) \equiv (p_{m,1}^L(x), \cdots, p_{m,j}(x), \cdots, p_{m,J}^L(x)),$$

$$\mathbf{a}_{m,-j}(x) \equiv (p_{m,1}(x), \cdots, p_{m,j-1}(x), p_{m,j+1}(x), \cdots, p_{m,J}(x)).$$

$\mathbf{a}_{m,j}^L(x)$ has lower bounds everywhere except its j-th element. $\mathbf{a}_{m,-j}(x)$ is a subvector of $\mathbf{a}_m(x)$ which omits $p_{m,j}(x)$.

Theorem 5. *Let Assumption 4-5 hold. For all $j \in \mathcal{J}$, $\delta_{0j}(x) = p_{1,j}^L(x)$. Given the lower bounds, $p_{m+1,j}^U(x)$ is the unique root of*

$$1 - \gamma_m(x) = V_C([\mathbf{a}_{m+1,j}^L(x), \mathbf{1}])$$

for all $m \geq 0$ and $j \in \mathcal{J}$. For $m \geq 1$ and all $j \in \mathcal{J}$, given the bounds for $p_{m,j}(x)$, the lower bounds $p_{m+1,j}^L(x)$ are derived from the following constrained minimization problem.

$$p_{m+1,j}^L(x) = \min_{\mathbf{a}_{m,-j}(x)} V_{Cj}^{-1}(\mathbf{a}_{m,-j}(x), 1 - \gamma_{m-1}(x) - \delta_{mj}(x))$$

$$s.t. \ V_C([\boldsymbol{a}_m(x), \boldsymbol{1}]) = 1 - \gamma_{m-1}(x) \quad and \quad p_{m,j}(x) \in [p_{m,j}^L(x), p_{m,j}^U(x)], \ \forall j \in \mathcal{J}.$$

Proof of Theorem 5. The lower bounds for $p_{1,j}(x)$ are trivial. They are directly obtained from the inequalities of (2.8) given that $p_{0,j}(x) = 0$ for all j and x. The equalities of (2.8),

$$1 - \gamma_m(x) = V_C([\mathbf{a}_{m+1}(x), \mathbf{1}]),$$

must hold for all j and $m \geq 0$. For the upper bounds, notice that the volume function V_C is decreasing in every element of $\mathbf{a}_{m+1}(x)$. Given the lower bounds of $\{p_{m+1,j}^L(x)\}_{j \in \mathcal{J}}$, the largest possible value of $p_{m+1,j}(x)$ is obtained from the equality when $p_{m+1,i}(x) = p_{m+1,i}^L(x)$ for all $i \neq j$. The solution is unique as the volume function is monotone in each argument of $\mathbf{a}_{m+1}(x)$.

To show the results for the lower bounds, the equalities of (2.8) can be rewritten as

$$1 - \gamma_{m-1}(x) = V_C([\mathbf{a}_m(x), \boldsymbol{1}_{m+1,j}(x)]) + V_C([(p_{m,1}(x), \cdots, p_{m+1,j}(x), \cdots, p_{m,J}(x)), \mathbf{1}])$$

Substituting the above equalities into the inequalities of (2.8) yields

$$V_C([(p_{m,1}(x), \cdots, p_{m+1,j}(x), \cdots, p_{m,J}(x)), \mathbf{1}]) \leq 1 - \gamma_{m-1}(x) - \delta_{mj}(x). \quad \text{(A.1)}$$

Given any $\mathbf{a}_{m,-j}(x)$, the lowest value of $p_{m+1,j}(x)$ is obtained when the equality of (A.1) holds. Let $\mathcal{P}_{m+1,j}(x)$ denote the set of values of $p_{m+1,j}(x)$ satisfying (A.1) with equality given all possible $\mathbf{a}_{m,-j}(x)$ under constraints such that $1 - \gamma_{m-1}(x) = V_C([\mathbf{a}_m(x), \mathbf{1}])$ and $p_{m,j}(x) \in [p_{m,j}^L(x), p_{m,j}^U(x)]$ for all $j \in \mathcal{J}$. Then $p_{m+1,j}^L(x)$ is the minimum of $\mathcal{P}_{m+1,j}(x)$. As the lowest possible value of $p_{m+1,j}(x)$ given $\mathbf{a}_{m,-j}(x)$ is given by the inverse function $V_{Cj}^{-1}(\mathbf{a}_{m,-j}(x), 1 - \gamma_{m-1}(x) - \delta_{mj}(x))$, $p_{m+1,j}^L(x)$ is also obtained by minimising V_{Cj}^{-1} given the constraints. Q.E.D.

Solutions for the lower bounds of $p_{m+1,j}(x)$ involve constrained minimization. However, in practice, those are easily approximated by fixing $\mathbf{a}_{m,-j}(x)$ at the lower bounds of its all elements except one (any i-th element) which is fixed at its upper bound. For example, suppose $J = 3$. The approximated lower bound of $p_{m+1,1}(x)$ is given by

$$V_{C1}^{-1}(\mathbf{a}_{m,-1}(x), 1 - \gamma_{m-1}(x) - \delta_{mj}(x))$$

subject to $\mathbf{a}_{m,-1}(x) = (p_{m,2}^L(x), p_{m,3}^U(x))$ or $\mathbf{a}_{m,-1}(x) = (p_{m,2}^U(x), p_{m,3}^L(x))$. This is because all $\mathbf{a}_{m,-1}(x)$ satisfying the constraints produces numerically very

similar results.

Suppose the copula C is of an Archimedean family so that $C(u) = \phi^{-1}(\sum_{j=1}^{J} \phi(u_j))$ where $\phi(\cdot)$ is the generator function which is continuous, strictly decreasing, and convex. Noticing that the object of minimization comes from the inequalities (A.1), the LHS can approximated by the first order Taylor series expansion of $V_C([\mathbf{a}_m(x), \mathbf{1}])$ subject to $V_C([\mathbf{a}_{m,j}(x), \mathbf{1}]) = 1 - r_{m-1}(x)$.

$$V_C([(p_{m,1}(x), \cdots, p_{m+1,j}(x), \cdots, p_{m,J}(x)), \mathbf{1}]) \approx$$
$$1 - \gamma_{m-1}(x) + (p_{m+1,j}(x) - p_{m,j}(x))V_C^j([\mathbf{a}_{m,j}(x), \mathbf{1}])$$

where V_C^j is the partial derivative of V_C w.r.t. the j-th argument. V_C^j is negative and only depends on $p_{m,j}(x)$ given the constraint. As V_C^j has the minimum at $p_{m,j}(x) = p_{m,j}^L(x)$, the lowest value of $p_{m+1,j}(x)$ derived from the approximation is

$$p_{m+1,j}^L(x) \approx p_{m,j}^L(x) - \frac{\delta_{mj}(x)}{V_C^j([\mathbf{a}_{m,j}(x), \mathbf{1}])} \quad s.t. \;\; p_{m,j}(x) = p_{m,j}^L(x)$$

which is the same over all $a_{m,j}(x)$ as far as the constraints are satisfied.

Proofs of main results

Proof of Proposition 3. The U-level set $\mathcal{U}(y; h)$ is a subset of the unit hypercube such that given $y = (m, j)$,

$$\mathcal{U}(y; h) \equiv \left\{ T^* : T_j^* \in [\tau_m, \tau_{m+1}] \;\wedge\; T_j^* < T_i^* \text{ for all } i \neq j \right\}.$$

Let U_h denote the support of $\mathcal{U}(y; h)$ and Q_h be the collection of all connected unions of elements of U_h. Define the subcollection $U_h^{\mathcal{S}}$ of U_h as $U_h^{\mathcal{S}} \equiv \{\mathcal{U} \in U_h : \mathcal{U} \subseteq \mathcal{S}\}$. Then for all $\mathcal{S} \in Q_h \backslash U_h$, $\cup_{\mathcal{A} \in U_h^{\mathcal{S}}} \mathcal{A} = \mathcal{S}$ and any element of U_h has measure zero intersection with other elements of U_h. Therefore, by Theorem 3 of CR17, U_h is the collection of core determining test sets. Then a set of equalities (2.4) derived for all possible outcome $y = (m, j)$ where $m \in \mathcal{T}$, $j \in \mathcal{J}$ characterises the sharp identified set of $\mathcal{F}$ by Corollary 2 of CR17.   Q.E.D.

Proof of Theorem 2. The proof of Theorem 5 applies. The equations for upper bounds are derived from the equalities of (2.8). By expanding the volume function,

$$1 - \gamma_m(x) = 1 - p_{m+1,1}(x) - p_{m+1,2}(x) + C(p_{m+1,1}(x), p_{m+1,2}(x)).$$

and rearranging this equation yields

$$\gamma_m(x) - p_{m+1,j}(x) = p_{m+1,i}(x) - C(p_{m+1,1}(x), p_{m+1,2}(x)).$$

Therefore, the largest value of $p_{m+1,i}$ is obtained at the lower bound of $p_{m+1,j}(x)$.

Likewise, the inequalities of (2.8) become (A.1). By expanding the volume function,

$$\gamma_{m-1}(x) + \delta_{m1}(x) - p_{m,2}(x) \le p_{m+1,1}(x) - C(p_{m+1,1}(x), p_{m,2}),$$

$$\gamma_{m-1}(x) + \delta_{m2}(x) - p_{m,1}(x) \le p_{m+1,2}(x) - C(p_{m,1}(x), p_{m+1,2}),$$

and $p_{m+1,i}(x)$ has the lowest value when the equality holds given $p_{m,j}(x)$. Let the equalities hold. Then $p_{m+1,i}(x)$ has the lowest value when $p_{m,i}(x) = p_{m,i}^U(x)$ since the partial derivatives of C w.r.t. its arguments are positive and smaller than 1. $\hfill$ Q.E.D.

Proof of Lemma 1. The upper bounds are derived by the direct application of Theorem 5. From the equalities of (2.8) and Assumption 6,

$$\prod_{i \ne j}[1 - p_{m,i}(x)] = \frac{1 - \gamma_{m-1}(x)}{1 - p_{m,j}(x)}, \quad \forall j \in \mathcal{J}$$

By substitutiong the above equation into the inequalities of (2.8),

$$\delta_{mj}(x) \le \frac{1 - \gamma_{m-1}(x)}{1 - p_{m,j}(x)}(p_{m+1,j}(x) - p_{m,j}(x)).$$

Then rearranging the above inequalities yields

$$p_{m,j}(x) + \frac{\delta_{mj}(x)}{1 - \gamma_{m-1}(x)}(1 - p_{m,j}(x)) \le p_{m+1,j}(x),$$

and the RHS has the lowest value when $p_{m,j}(x) = p_{m,j}^L(x)$ since $\frac{\delta_{mj}(x)}{1-\gamma_{m-1}(x)} < 1$. $\hfill$ Q.E.D.

Proof of Lemma 2. In the first inequality of (2.8), the full expansion of RHS yields

$$V_C([\mathbf{a}_m^L(x), \mathbf{b}_{m+1,j}(x)]) \le p_{m+1}(x) - p_m(x)$$

and therefore,

$$\delta_{mj}(x) + p_m(x) \le p_{m+1}(x).$$

The above inequality needs to be satisfied for any possible values of $p_m(x)$. Thus, the lower bound is derived at the lowest possible value of $p_m(x)$.

For the upper bound, the expansion the LHS of the second equality of (2.8) followed by the Fréchet-Hoeffding bounds inequality yields the desired result.

$$p_{m+1,j}(x) \leq \sum_{j=1}^{J} \sum_{t=0}^{m} \delta_{tj}(x).$$

These bounds are identical to those in Peterson (1976) which are sharp.

<div align="right">Q.E.D.</div>

Proof of Theorem 3. Given the definition of the $\mathcal{U}$-level set $\mathcal{U}(m, j; h)$, let U_h denote the support of $\mathcal{U}(m, j; h)$ and Q_h be the collection of all connected unions of elements of U_h. Now suppose that the collection of core determining test sets, $\tilde{Q}_h$, is

$$\tilde{Q}_h \equiv \{\bigcup_{j \in \mathcal{N}} \mathcal{U}(m, j; h) : m \in \mathcal{T}, \mathcal{N} \subseteq \mathcal{J}\}.$$

Then $\tilde{Q}_h$ includes all possible unions the support of the $\mathcal{U}$-level set given $m \in \mathcal{T}$. For any $m \neq m'$ where $m, m \in \mathcal{T}$ and any $\mathcal{N}, \mathcal{N}' \subseteq \mathcal{J}$, the intersection

$$\bigcup_{j \in \mathcal{N}} \mathcal{U}(m, j; h) \cap \bigcup_{j \in \mathcal{N}'} \mathcal{U}(m', j; h)$$

has zero probability measure.

Suppose $\mathcal{S}$ is a connected union of $\mathcal{S}_1 \equiv \bigcup_{j \in \mathcal{N}} \mathcal{U}(m, j; h)$ and $\mathcal{S}_2 \equiv \bigcup_{j \in \mathcal{N}'} \mathcal{U}(m', j; h)$ where m and m' are consecutive integers. Then it is not core determining by Theorem 3 of CR17 because $\mathcal{S}_1$ and $\mathcal{S}_2$ has measure zero intersection. Any other connected unions $\tilde{\mathcal{S}} \equiv \mathcal{S} \cup \bigcup_{j \in \mathcal{N}''} \mathcal{U}(m'', j; h)$ is not core determining by the same logic. In this way, all the unions in $Q_h / \tilde{Q}_h$ are shown to be not core determining.

For the unions $\bigcup_{j \in \mathcal{J}} \mathcal{U}(m, j; h)$, either $\mathcal{U}(y; h) \subseteq \bigcup_{j \in \mathcal{J}} \mathcal{U}(y; h)$ or $\mathcal{U}(y; h) \subseteq cl(\bigcup_{j \in \mathcal{J}} \mathcal{U}(y; h))^c$ is satisfied for all possible outcomes $y = (m, j)$. Therefore, these sets produce moment equalities by Corollary 2 of CR17.   Q.E.D.

Proof of Theorem 4. By Assumption 9, $p_{1,1}^{L} = \sup_{z \in \mathcal{R}_Z} p_{1,1}^{L}(z)$ where $p_{1,1}^{L}(z) = \delta_{01}(z)$ and $p_{1,2}^{L}(z) = p_{1,2}^{L}(z)$. Given the lower bounds of $p_{m,1}$ and $p_{m,2}(z)$, the upper bounds of $p_{1,1}(z)$ and $p_{1,2}(z)$ are given by Theorem 2. Then, $p_{m+1,1}^{U} = \inf_{z \in \mathcal{R}_Z} p_{m+1,1}^{U}(z)$. Given the upper and lower bounds of $p_{m,1}$ and $p_{m,2}(z)$, the lower bounds of $p_{m,1}(z)$ and $p_{m,2}(z)$ are given by Theorem 2. Then, $p_{m+1,1}^{U} =$

$\inf_{z \in \mathcal{R}_Z} p^U_{m+1,1}(z).$ Q.E.D.

Bibliography

ABBRING, J. H. and VAN DEN BERG, G. J. (2003). The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65** (3), 701–710.

ANDREWS, D. W. and BARWICK, P. J. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, **80** (6), 2805–2826.

— and GUGGENBERGER, P. (2009). Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. *Econometric Theory*, **25** (3), 669–709.

— and HAN, S. (2009). Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities. *The Econometrics Journal*, **12** (s1).

— and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** (1), 119–157.

ANGRIST, J. D. and EVANS, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, **88** (3), 450–477.

— and KRUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, **106** (4), 979–1014.

BERRINGTON, A. and DIAMOND, I. (2000). Marriage or cohabitation: A competing risks analysis of first-partnership formation among the 1958 british birth cohort. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **163** (2), 127–151.

BLAES, A., PRIZMENT, A., KOENE, R. J. and KONETY, S. (2017). Cardio-oncology related to heart failure: common risk factors between cancer and cardiovascular disease. *Heart failure clinics*, **13** (2), 367–380.

BOOTH, A. L. and SATCHELL, S. E. (1995). The hazards of doing a phd: an analysis of completion and withdrawal rates of british phd students in the 1980s. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 297–318.

BUGNI, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, **78** (2), 735–753.

—, CANAY, I. A. and SHI, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, **8** (1), 1–38.

CAMERON, A. C. and TRIVEDI, P. K. (2013). *Regression analysis of count data*, vol. 53. Cambridge university press.

CARD, D., JOHNSTON, A., LEUNG, P., MAS, A. and PEI, Z. (2015). The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in missouri, 2003-2013. *American Economic Review*, **105** (5), 126–30.

CARRIÈRE, J. F. (1995). Removing cancer when it is correlated with other causes of death. *Biometrical Journal*, **37** (3), 339–350.

CHEN, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72** (2), 235–251.

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** (5), 1243–1284.

—, LEE, S. and ROSEN, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, **81** (2), 667–737.

CHESHER, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, **73** (5), 1525–1550.

— (2010). Instrumental Variable Models for Discrete Outcomes. *Econometrica*, **78** (2), 575–601.

— and ROSEN, A. M. (2017). Generalized instrumental variable models. *Econometrica*, **85** (3), 959–989.

— and SMOLINSKI, K. (2012). Iv models of ordered choice. *Journal of Econometrics*, **166** (1), 33–48.

CIA (2018). *The world factbook 2018-2019*. Skyhorse Publishing.

COX, D. R. (1962). *Renewal theory*, vol. 1. Methuen London.

CUTLER, D. M. (2008). Are we finally winning the war on cancer? *Journal of Economic Perspectives*, **22** (4), 3–26.

DENG, Y., QUIGLEY, J. M. and ORDER, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, **68** (2), 275–307.

DRIVER, J. A., DJOUSSÉ, L., LOGROSCINO, G., GAZIANO, J. M. and KURTH, T. (2008). Incidence of cardiovascular disease and cancer in advanced age: prospective cohort study. *Bmj*, **337**, a2467.

DUARTE, C. W., LINDNER, V., FRANCIS, S. A. and SCHOORMANS, D. (2017). Visualization of cancer and cardiovascular disease co-occurrence with network methods. *JCO Clinical Cancer Informatics*, **1**, 1–12.

FALLICK, B. and RYU, K. (2007). The recall and new job search of laid-off workers: a bivariate proportional hazard model with unobserved heterogeneity. *The Review of Economics and Statistics*, **89** (2), 313–323.

FAN, Y. and LIU, R. (2018). Partial identification and inference in censored quantile regression. *Journal of Econometrics*.

FANG, F., FALL, K., MITTLEMAN, M. A., SPARÉN, P., YE, W., ADAMI, H.-O. and VALDIMARSDÓTTIR, U. (2012). Suicide and cardiovascular death after a cancer diagnosis. *New England Journal of Medicine*, **366** (14), 1310–1318.

FARBER, H. S., ROTHSTEIN, J. and VALLETTA, R. G. (2015). The effect of extended unemployment insurance benefits: Evidence from the 2012-2013 phase-out. *American Economic Review*, **105** (5), 171–76.

— and VALLETTA, R. G. (2015). Do extended unemployment benefits lengthen unemployment spells? evidence from recent cycles in the us labor market. *Journal of Human Resources*, **50** (4), 873–909.

FLINN, C. and HECKMAN, J. (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, **18** (1), 115–168.

GALICHON, A. and HENRY, M. (2011). Set identification in models with multiple equilibria. *The Review of Economic Studies*, **78** (4), 1264–1298.

GREGORY-SMITH, I., THOMPSON, S. and WRIGHT, P. W. (2009). Fired or retired? a competing risks analysis of chief executive turnover. *The Economic Journal*, **119** (536), 463–481.

HAN, A. and HAUSMAN, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of applied Econometrics*, **5** (1), 1–28.

HECKMAN, J. J. and HONORÉ, B. E. (1989). The identifiability of the competing risks model. *Biometrika*, pp. 325–330.

HO, K. and ROSEN, A. M. (2015). *Partial identification in applied research: benefits and challenges.* Tech. rep., National Bureau of Economic Research.

HONORÉ, B. E. and LLERAS-MUNEY, A. (2006). Bounds in competing risks models and the war on cancer. *Econometrica*, **74** (6), 1675–1698.

HOROWITZ, J. L. and MANSKI, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American statistical Association*, **95** (449), 77–84.

HUANG, X. and ZHANG, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics*, **64** (4), 1090–1099.

IMBENS, G. W. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, **72** (6), 1845–1857.

JOE, H. (1997). *Multivariate models and multivariate dependence concepts.* Chapman and Hall/CRC.

KAIDO, H., MOLINARI, F. and STOYE, J. (2017). Confidence intervals for projections of partially identified parameters. *CeMMAP Working Papers*, **CWP47/17**.

KATZ, L. F. and MEYER, B. D. (1990). The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of public economics*, **41** (1), 45–72.

KLINE, B. and TAMER, E. (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics*, **7** (2), 329–366.

KOENE, R. J., PRIZMENT, A. E., BLAES, A. and KONETY, S. H. (2016). Shared risk factors in cardiovascular disease and cancer. *Circulation*, **133** (11), 1104.

KOENKER, R. (2005). *Quantile regression*. 38, Cambridge university press.

KROFT, K., LANGE, F. and NOTOWIDIGDO, M. J. (2013). Duration dependence and labor market conditions: Evidence from a field experiment. *The Quarterly Journal of Economics*, **128** (3), 1123–1167.

—, —, — and KATZ, L. F. (2016). Long-term unemployment and the great recession: the role of composition, duration dependence, and nonparticipation. *Journal of Labor Economics*, **34** (S1), S7–S54.

LEE, S. (2006). Identification of a competing risks model with unknown transformations of latent failure times. *Biometrika*, pp. 996–1002.

— and LEWBEL, A. (2013). Nonparametric identification of accelerated failure time competing risks models. *Econometric Theory*, **29** (5), 905–919.

LIU, R. (2016). *A Competing Risks Model with Time-varying Heterogeneity and Simultaneous Failure*. Tech. rep., Department of Economics, Emory University (Atlanta).

LO, S. and WILKE, R. A. (2010). A copula model for dependent competing risks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59** (2), 359–376.

LO, S. M. and WILKE, R. A. (2016). Identifiability of the sign of covariate effects in the competing risks model. *Econometric Theory*, pp. 1–32.

LOMBORG, B. (2003). *The skeptical environmentalist: measuring the real state of the world*, vol. 1. Cambridge University Press Cambridge.

MANSKI, C. and PEPPER, J. (2009). More on monotone instrumental variables. *The Econometrics Journal*, **12** (S1), S200–S216.

MANSKI, C. F. and PEPPER, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, **68** (4), 997–1010.

McCALL, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica: Journal of the Econometric Society*, pp. 647–682.

MOSLEHI, J. J. (2016). Cardiovascular toxic effects of targeted cancer therapies. *New England Journal of Medicine*, **375** (15), 1457–1467.

MULLAHY, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics*, **79** (4), 586–593.

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

NICKELL, S. (1979). Education and lifetime patterns of unemployment. *Journal of Political Economy*, **87** (5, Part 2), S117–S131.

PETERSON, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences*, **73** (1), 11–13.

POPE III, C. A., BURNETT, R. T., TURNER, M. C., COHEN, A., KREWSKI, D., JERRETT, M., GAPSTUR, S. M. and THUN, M. J. (2011). Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships. *Environmental health perspectives*, **119** (11), 1616.

RIVEST, L.-P. and WELLS, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis*, **79** (1), 138–155.

ROMANO, J. P. and SHAIKH, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica*, **78** (1), 169–211.

ROSEN, A. M. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics*, **146** (1), 107–117.

SEER (2017). SEER cancer statistics review, 1975-2014. *National Cancer Institute*.

STOYE, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, **77** (4), 1299–1315.

SUEYOSHI, G. T. (1992). Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of econometrics*, **51** (1-2), 25–58.

TERZA, J. V., BASU, A. and RATHOUZ, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, **27** (3), 531–543.

TSIATIS, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, **72** (1), 20–22.

WHEELOCK, D. C. and WILSON, P. W. (2000). Why do banks disappear? the determinants of us bank failures and acquisitions. *Review of Economics and Statistics*, **82** (1), 127–138.

WINDMEIJER, F. A. and SANTOS SILVA, J. M. (1997). Endogeneity in count data models: an application to demand for health care. *Journal of applied econometrics*, **12** (3), 281–294.

WOOLDRIDGE, J. M. (1992). Some alternatives to the box-cox regression model. *International Economic Review*, pp. 935–955.

ZHENG, M. and KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82** (1), 127–138.