A Grouped Mixed Proportional Hazard Model with Social Interactions: The Passage of the Motorcycle-Helmet-Use Law

JONGYEARN LEE [*]   YOONSEOK LEE [†]

September 2011

Abstract

We develop a mixed proportional hazard model in discrete time when there is cross-sectional duration dependence from social interactions. We model the cross-sectional dependence using the weighted lagged choices of neighbors based on a proper spatial weight matrix, and non-parametrically specify the baseline hazard and the distribution of unobserved heterogeneity. We use EM algorithm to estimate the duration model and derive the observed information matrix for statistical inference. Using the U.S. state-level panel data, we analyze that the state legislation decision on the mandatory motorcycle-helmet-use law is significantly affected by neighboring states' choices, whereas the fatality rate from motorcycle-related accidents is not so.

Keywords: Mixed proportional hazard, social interaction, cross section dependence, legislative decision making, motorcycle-helmet-use law.

JEL Classifications: C23, C41, K32.

---

[*]Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109; *email*: `yearn@umich.edu`

[†]Corresponding author. Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109; *email*: `yoolee@umich.edu`

1 Introduction

The main purpose of this paper is to understand the decision making mechanism of state legislations and, in particular, to find an evidence that coverage of the motorcycle-helmet-use law (MHU law, hereafter) in a state is spatially dependent on neighboring states' status or decisions.

To analyze such a social interaction aspect in state legislation decision making, we specify a discrete choice model with social interactions as Brock and Durlauf (2001a, 2001b), which incorporates the social interaction term in the random utility maximization problem. Individual expectation to others' decision is the key component deriving the social interactions. In this particular example, however, each decision making can be only realized at certain times, such as only when there is a legislate meeting (i.e., timing friction in decision realization). By introducing such timing friction in the discrete choice model, similarly as the job searching model of Lancaster (1979), we derive a hazard model with duration dependence. In particular, we introduce a grouped mixed proportional hazard (MPH) model (e.g., Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978) with social interactions, where the hazard rate is a function of other states' discrete choices (e.g., Carruthers, Guinnane and Lee, 2010). The baseline hazard and the distribution of the unobserved heterogeneity are specified nonparametrically (e.g., Prentice and Gloeckler, 1978; Heckman and Singer, 1984; Meyer, 1990, 1995). The EM-algorithm is used to facilitate the estimation of the MPH model with social interactions (e.g., Lee, 2007).

Estimation results using U.S. state-level panel data from 1975 to 2006 show statistically significant interactions with neighboring states' decisions on the MHU law, whereas safety concern is found not to be important when the policy makers make decisions. Though this analysis does not give an answer to an issue whether introducing the mandatory MHU law is beneficial or not, it explains a behavioral aspect of the legislative decision making procedure in the context of social interactions and empirically shows how the proximity between agents affects the decision making.

The remainder of the paper is organized as follows. Section 2 describes the main model, discrete choice with social interactions and timing friction in decision realization. Section 3 develops a mixed proportional hazard model with grouped data, that includes both the interac-

tion term and the unobserved heterogeneity, and provides econometric foundation of the choice model. Section 4 introduces the EM algorithm as an estimation method. Section 5 delineates the data and discusses the estimated results. Section 6 concludes the paper with some remarks. Technical details are provided in the Appendix.

2 Discrete Choice with Social Interactions

2.1 Random utility maximization

We consider individuals $i = 1, \cdots, n$ who can switch their choices over time $t = 1, \cdots, T$. The binary choice is denoted by an indicator variable $d_i(t)$, which has support $\{-1, 1\}$. The observed characteristics of each individual i, possibly time-varying, are denoted as a $g \times 1$ vector $x_i(t)$. The unobservable independent random private utility (or a random shock) is denoted by $\varepsilon_i(t, d_i(t))$, which depends on the realized individual's choice and is independent of $x_i(t)$ for all i and t.

Similarly as Brock and Durlauf (2001a, 2001b), we impose social interactions in the individual decision by assuming that the expected behavior of others influence each individual choice. Examples are spill-over effects, externalities and peer-effects. More precisely, we assume a twice-differentiable instantaneous individual utility function given by $V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t)))$, where $d_{-i}(t) = (d_1(t), \cdots, d_{i-1}(t), d_{i+1}(t), \cdots, d_n(t))$ denotes the vector of choices other than that of individual i and $\mu_i^e(d_{-i}(t))$ represents individual's belief concerning the choices of other agents. Given that individuals are myopic so that they only make choices by comparing current utilities without considering future paths of choices, each choice is described by solving

$$\max_{d_i(t) \in \{-1,1\}} V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) \tag{1}$$

for each t. Note that myopic behavior can be understood in the context of repeated search or infinite discount rate, and it well justifies the proportional hazard specification (e.g., van den Berg, 2001).

We assume that the individual utility function can be represented as

$$V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) = u(d_i(t), x_i(t)) + \varepsilon_i(t, d_i(t)) \qquad (2)$$
$$+ U_S(d_i(t), x_i(t), \mu_i^e(d_{-i}(t))),$$

where $u(d_i(t), x_i(t))$ and $U_S(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)))$ are observable deterministic private and social utilities, respectively. Without the social utility U_S, (2) corresponds to the standard random utility function. We further let the deterministic private utility be linear as (e.g., Brock and Durlauf, 2001b, p.3307)

$$u(d_i(t), x_i(t)) = d_i(t)u_1(x_i(t)) + u_2(x_i(t)), \qquad (3)$$

which is without loss of generality since it coincides with the original utility function on the support of the individual choices $\{-1, 1\}$. The social utility possesses a generalized quadratic conformity effect as

$$U_S(d_i(t), x_i(t), \mu_i^e(d_{-i}(t))) = -\frac{\alpha}{2} \mathbb{E}_{i,t} \left[\sum_{j \neq i} w_{ij}(t)(d_i(t) - d_j(t))^2 \right], \qquad (4)$$

where α is an unknown scalar parameter and $\mathbb{E}_{i,t}[\cdot]$ denotes the conditional expectation of i at t given the values of $(d_i(t), x_i(t))$. In this specification, $\alpha w_{ij}(t) = \partial^2 V / \partial d_i(t) \partial \mathbb{E}_{i,t}[d_j(t)] = \partial^2 U_S / \partial d_i(t) \partial \mathbb{E}_{i,t}[d_j(t)]$ measures the strategic complementarity between individual choices and the expected choices of others (e.g., Cooper and John, 1988; Brock and Durlauf, 2001b). The term $w_{ij}(t)$ in (4) represents the interaction weight between individuals i and j. For each t, we define $w_{ij}(t) = M(||\varphi_i(t), \varphi_j(t)||)$ for $i, j = 1, 2, \ldots, n$, where $w_{ii}(t) = 0$, $w_{ij}(t) = w_{ji}(t)$ and $||\cdot, \cdot||$ is a proxy of economic distance between i and j. More precisely, $||\cdot, \cdot||$ is a distance function of a pair of characteristics $\varphi_i(t)$ and $\varphi_j(t)$, and $M(\cdot)$ is a nonnegative and strictly increasing function with $M(0) = 0$. Note that $w_{ij}(t)$ are typically assumed to be fixed and deterministic for the identification purposes (e.g., Manski, 1993). We implicitly assume that there is no individual without a neighbor; the choice of neighborhood is also fixed and not endogenous.

For $d_i^2(t) = 1$ we can rewrite (4) as $\alpha \sum_{j \neq i} w_{ij}(t)(d_i(t)\mathbb{E}_{i,t}[d_j(t)] - 1)$, and obtain the values

3

of expectations $\mathbb{E}_{i,t}[d_j(t)]$ by assuming that all individuals have rational expectations for each t:

$$\mathbb{E}_{i,t}\left[d_j(t)\right] = \mathbb{E}_t\left[d_j(t)|x_1(t),\cdots,x_n(t);\mathbb{E}_{k,t}d_\ell(t) \text{ for } k,\ell = 1,\cdots,n\right]. \tag{5}$$

The solutions satisfying this self-consistent condition (e.g., Brock and Durlauf, 2001a) close the model, provided they exist. Uniqueness of the self-consistent equilibrium gives the identification condition in this framework. More precisely, from the standard random utility maximization, individual i chooses $d_i(t)$ over $-d_i(t)$ at time t if

$$V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) \geq V(-d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, -d_i(t))), \tag{6}$$

whose probability is given by

$$\mathbb{P}\left\{V(d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t))) \geq V(-d_i(t), x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, -d_i(t)))\right\}$$
$$= \mathbb{P}\left\{\varepsilon_i(t, -d_i(t)) - \varepsilon_i(t, d_i(t)) \leq 2\left[u_1(x_i(t)) + \alpha\sum\nolimits_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]\right]d_i(t)\right\}$$
$$= G\left(2\left[u_1(x_i(t)) + \alpha\sum\nolimits_{j\neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]\right]d_i(t)\right), \tag{7}$$

where we assume that $\varepsilon_i(t, -1) - \varepsilon_i(t, 1)$ is independent and identically distributed (i.i.d.) over i and t with $G(\cdot)$ being the distribution function that is symmetric around zero. Using this result, the self-consistent solution $\overline{\mu}(t) = \mathbb{E}_{i,t}[d_j(t)]$ satisfying (5) can be obtained by solving

$$\overline{\mu}(t) = H\left(u_1(x_i(t)) + \alpha\sum\nolimits_{j\neq i} w_{ij}(t)\overline{\mu}(t)\right), \tag{8}$$

which is assumed to exist uniquely, where $H(y) = 2G(2y) - 1$ for some $y \in \mathbb{R}$.

Assumption 1 (self-consistent equilibrium) *Given $H(\cdot)$, $u_1(x_i(t))$ and $\alpha\sum_{j\neq i} w_{ij}(t)$, there exists a unique self-consistent expectation $\overline{\mu}(t)$ satisfying (8) for each t.*

Note that the unique existence of the self-consistent equilibrium requires assumptions on the distribution $G(\cdot)$ as well as $u_1(x_i(t))$ and $\alpha\sum_{j\neq i} w_{ij}(t)$. For example, when G is logistic, $H(\cdot) = \tanh(\cdot)$ and thus the existence of self-consistent equilibrium in (8) follows immediately (e.g.,

Brock and Durlauf, 2001a, Proposition 1). Moreover, provided that $\alpha \sum_{j \neq i} w_{ij}(t) \leq 1$, which holds if $\alpha \leq 1$ under the row normalization assumption (i.e., $\sum_{j \neq i} w_{ij}(t) = 1$ for each i and t) in Assumption 6, the equilibrium is unique from the properties of the $\tanh(\cdot)$ function. When $\alpha \sum_{j \neq i} w_{ij}(t) > 1$, on the other hand, the uniqueness can be obtained only when $|u_1(x_i(t))|$ is large enough (e.g., Brock and Durlauf, 2001a, Proposition 2), which is the case of the estimation result in Section 5.

2.2 Choice with timing frictions

Though we introduce the social interaction term in the choice mechanism, the random utility analysis in the previous subsection is rather standard since we assume that each individual is myopic and the utility maximization problem is solved for each t. There could be the cases, however, that the choice is not allowed for some t even though the random utility maximization tells so. A passage of a law is a good example: Let the agent i be the state legislature and the passage of a particular law is determined by solving the random utility maximization. For some cases, even when people want to pass the law soon, the number of legislate meetings is limited and thus there could be an exogenous timing friction in realizing the choice.

To incorporate such an idea into the framework, we suppose that an alarm clock is assigned to each individual i, where the alarms are independent over time. When the clock rings, the agent has an opportunity to revise her choice. The choice and the alarm are mutually independent, and the choice is assumed to be made when the clock rings or right before at $t^- = \lim_{\Delta \to 0}(t - \Delta)$ for $\Delta > 0$. More precisely, we let the occurrence of alarm follows the time-dependent (or non-homogeneous) Poisson process with rate $\rho_i(t) > 0$, so that the expected number of alarms in the time interval $[0, t)$ is given as $\int_0^t \rho_i(s)ds$ for each i. We further impose a common factor structure on $\rho_i(t)$ as

$$\rho_i(t) = \lambda_0(t)v_i \tag{9}$$

with $\lambda_0(t) > 0$ and $v_i > 0$, where $\lambda_0(t)$ is the common time-dependent rate across individuals while the time-invariant (unobservable) heterogeneity v_i allows for variations over i.

If individual i followed the decision rule in (6), her revision of the choice is observed in the

short time period $[t, t + \Delta)$ with $\Delta > 0$ if and only if (i) her alarm clock rings at t and (ii) the utility with the new choice exceeds that with the old one. Apparently, the probability of the events (i) and (ii) are $\rho_i(t)\Delta = \lambda_0(t)v_i\Delta$ and $G(2[u_1(x_i(t)) + \alpha \sum_{j \neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]]d_i(t))$, respectively from (9) and (7). Since these events are assumed to be mutually independent, the probability of choice revision in $[t, t + \Delta)$ conditional on no revision occurred before t is given by their product and it yields the *hazard function* for the choice change of individual i:

$$\lambda_0(t)G\left(2[u_1(x_i(t)) + \alpha \sum_{j \neq i} w_{ij}(t)\mathbb{E}_{i,t}[d_j(t)]]d_i(t)\right)v_i, \tag{10}$$

which is in the form of the mixed proportional hazard (MPH) model with the baseline hazard function $\lambda_0(t)$ and the unobservable heterogeneity v_i (e.g., Lancaster, 1990; van den Berg, 2001).

We remark that knowing the characteristic of the choice-change-allowance process is crucial to correctly identify the choice behavior. For example, suppose we ignore the choice-change-allowance process and simply observe the choice behaviors at a fixed frequency. Then for any two identical consecutive choices of individual i, say $d_i(t) = d_i(t') = 1$ for $t < t'$, we cannot tell which scenario results in such observations among the followings: (i) $d_i(t') = 1$ because $V_i(t', 1) - V_i(t', -1) > 0$, where $V_i(t, d_i(t)) = V(1, x_i(t), \mu_i^e(d_{-i}(t)), \varepsilon_i(t, d_i(t)))$, and the choice revision is allowed at t'; (ii) $d_i(t') = 1$ because the choice revision is not allowed at t' whether the sign of $V_i(t', 1) - V_i(t', -1)$ is positive or negative. Particularly when $V_i(t', 1) - V_i(t', -1) < 0$ but the individual cannot change her choice because it is not allowed at t', it violates the fundamentals of the standard random utility maximization problem.

3 Grouped MPH Models with Social Interactions

3.1 Semiparametric duration models with grouped data

Though the failure time is continuous, the standard panel data only provide observations on failure times aggregated up to discrete intervals (i.e., grouped duration data; e.g., Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978). To handle this discrepancy, we suppose failure times are grouped into intervals $B_s = [b_{s-1}, b_s)$ for $s = 1, 2, \cdots, S$ with $b_0 = 0$ and

$b_S = \infty$ without loss of generality, where the length of each interval corresponds to the panel survey frequency. Survival to time b_s is the same as surviving until the s-th interval B_s (e.g., Sueyoshi, 1995) and the failure time of individual i in B_s are recorded as $\tau_i = s$. Since we usually deal with equi-spaced panel data, we simply let $b_t = t$ for all $t = 1, \cdots, T$ and consider T intervals: $[0, 1), [1, 2), \ldots, [T - 1, T)$ ignoring the last interval $[T, \infty)$ that is after the survey period and thus all durations lasting over T are naturally right censored. We further assume that covariates are at best recorded up to intervals and the values do not change in each interval $[t - 1, t)$. Following the standard notations of panel data, we simply rewrite $x_i(t)$, $w_{ij}(t)$, $d_i(t)$ as $x_{i,t}$, $w_{ij,t}$, $d_{i,t}$, respectively, in what follows.

In order to make the empirical analysis tractable, we further impose three more assumptions. First, we specify $u_1(x_{i,t}) = x'_{i,t}\beta$ for a $g \times 1$ parameter vector β as the standard discrete choice or the Cox's (1972) MPH models. Second, the interaction weight $w_{ij,t}$ is time invariant so that it is simply denoted as w_{ij}. Since the main analysis is based on the geographical proximity as a measure of the interaction weights, this assumption holds naturally. However, as long as $w_{ij,t}$ is predetermined, time varying weight can be considered. Third, knowing that the renewal of the choice is not highly frequent in the data set (only 42 revisions are made over 383 time periods), we assume that the expectation for neighbors' current choices is equal to the previous choices (i.e., $\mathbb{E}_{i,t}[d_{j,t}] = d_{j,t-1}$ for all $j \neq i$ and t) similarly as Wallis (1980). Using the grouped duration, we thus define the hazard rate as

$$\lambda_i \left(t | x_{i,t}, \sum_{j \neq i} w_{ij} d_{j,t-1}, v_i \right) = \lambda_0(t) \phi \left(x'_{i,t}\beta + \alpha \sum_{j \neq i} w_{ij} d_{j,t-1} \right) v_i \tag{11}$$

from (10) for all $i = 1, \cdots, n$ and $t = 1, \cdots, T$, where $\phi(y) = G(2y)$ if the hazard is from the choice -1 to 1, and $\phi(y) = 1 - G(2y)$ if it is from 1 to -1.

Heckman and Singer (1984) suggest that the distribution of the unobserved heterogeneity v_i be nonparametrically estimated in order to avoid any misspecification problem; as the number of mass points increases, discrete distributions can approximate any distribution arbitrarily well. The nonparametric estimator, however, is very sensitive to the assumed shape of the baseline hazard function $\lambda_0(t)$ (e.g., Trussell and Richards, 1985) especially with single-spell data. For

7

a possible solution, Meyer (1990, 1995) proposes to use piecewise constant baseline hazard functions (e.g., Prentice and Gloeckler, 1978) as well as the Heckman-Singer approach. More precisely, we let the baseline hazard $\lambda_0(t)$ be piecewise constant:

$$\lambda_0(t) = \exp\left(\gamma_\ell\right) \quad \text{if } t \in [a_{\ell-1}, a_\ell), \tag{12}$$

where $1 = a_1 < \cdots < a_h = T$ (with $h \leq T$) is a subsequence of $t = 1, \cdots, T$. Such specification is useful especially when the hazard rate has much fluctuation or frequent peaks. It extracts common deterministic time trends from the covariates as the standard time effects in panel regressions. Note that γ_ℓ satisfies $\log \int_{a_{\ell-1}}^{a_\ell} \lambda_0\left(r\right) dr = \gamma_\ell$ for the continuous time case and the most flexible case is to let $a_t = t$ so that $\gamma_t = \log \int_{t-1}^{t} \lambda_0\left(r\right) dr$ for all t. For the unobserved heterogeneity, we assume v_i to be *i.i.d.* with a discrete distribution with m supports, whose density is given as

$$f\left(v\right) = \sum\nolimits_{j=1}^{m} p_j 1\left\{v = q_j\right\}, \tag{13}$$

where $\mathbb{E}v_i = 1$, $\sum_{j=1}^{m} p_j = 1$, $0 < p_j < 1$ and $0 < q_j < \infty$ for all $j = 1, 2, \cdots, m$. $1\left\{\cdot\right\}$ is the binary indicator.

3.2 Regularity conditions

Based on the specification (11), (12) and (13) in the previous section, we consider an MPH model given by

$$\lambda_i\left(t|v_i\right) = \exp(\gamma_t)\phi\left(x_{i,t}'\beta + \alpha \sum\nolimits_{j\neq i} w_{ij}d_{j,t-1}\right) v_i \tag{14}$$

for some known function $\phi : \mathbb{R} \to \mathbb{R}_+$, where v_i is *i.i.d.* with density given in (13) and independent of $z_{i,t} = (x_{i,t}', \sum_{j\neq i} w_{ij}d_{j,t-1})'$. Carruthers, Guinnane and Lee (2010) use a similar method with the exponential link function as a particular example of $\phi(\cdot)$. The weighted sum of $d_{j,t-1}$ by the interaction weights w_{ij} can be interpreted as the average influence of other agents' past decisions on i that can be understood as the individual i's expectation on the others' behavior or a learning effect. We first assume the following conditions.

Assumption 2 (failure time) *The failure time $\tau_i > 0$ is independent across i conditional on $z_{i,t}$ and v_i for each t; τ_i is independent of the censoring time C_i for all i.*

Assumption 3 (unobserved heterogeneity) *The unobserved heterogeneity $v_i > 0$ is i.i.d. of finite mixture (13) with $\mathbb{E}v_i = 1$; v_i is independent of $z_{i,t}$ and C_i for all i and t.*

Assumption 4 (baseline hazard) *The baseline hazard $\lambda_0(t)$ is nonnegative and piecewise constant given by (12) with $a_t = t$ for all t.*

Heckman and Singer (1984) assume that the distribution of the censoring variable C_i is known and independent of the covariate $z_{i,t}$ to show the consistency of maximum likelihood estimators with nonparametric unobserved heterogeneity v_i. In our case, the censoring only occurs at the fixed time T when the panel survey is over, so those assumptions hold naturally. $\mathbb{E}v_i$ is usually normalized to one so that the expected hazard rate becomes the unconditional hazard rate with no unobserved heterogeneity: $\mathbb{E}[\lambda_i(t|v_i)|z_{i,t}] = \lambda_i(t)$, where $\lambda_i(t) = \exp(\gamma_t)\phi(x'_{i,t}\beta + \alpha\sum_{j\neq i} w_{ij}d_{j,t-1})$. Surely the finite mean condition of v_i restricts its tail behavior and it is necessary for identification (e.g., van den Berg, 2001). The number of support points of v is assumed to be finite and fixed, though identification can be obtained even with increasing number of supports (e.g., Kiefer and Wolfowitz, 1956; Heckman and Singer, 1984; Meyer, 1995). We assume further conditions for identifying λ_0, β, α, and the distribution of v.

Assumption 5 (covariates) *(i) $z_{i,t} \in \mathcal{Z}$ for an open set $\mathcal{Z}$ in $\mathbb{R}^{g+1}$ and $Z = (z_{1,1}, \cdots, z_{n,T})'$ is of full column rank. (ii) No element of $x_{i,t}$ is constant and at least one argument of $x_{i,t}$ is defined on the continuum. (iii) The regression function ϕ is nonlinear and differentiable on $\mathcal{Z}$.*

Ignoring the social interaction term, Assumptions 3, 4, and 5 yield identifiably of our model (14) up to a constant multiplication as Elbers and Ridder (1982). In particular, Assumption 3 is the same as Assumption 1 of Elbers-Ridder; Assumption 4 satisfies Assumption 2 of Elbers-Ridder because $\int_0^t \lambda_0(r)dr = \sum_{s=1}^t \exp(\gamma_s)$ is an increasing function of $t \geq 0$; Assumption 5 and the form of the proportional hazard function satisfy Assumption 3 of Elbers-Ridder. Adding the social interaction term $\sum_{j\neq i} w_{ij}d_{j,t-1}$ in $z_{i,t}$ does not change the identification result as long as

Z is of full column rank and ϕ is nonlinear in $z_{i,t}$ (e.g., Brock and Durlauf, 2001b). Note that the duration model we consider here does not depend on the durations of others directly; instead the durations are dependent indirectly by including the choice-dependent term $\alpha \sum_{j \neq i} w_{ij} d_{j,t-1}$ in the hazard rate. The identification in this case, therefore, can be obtained as the standard MPH models by considering $\alpha \sum_{j \neq i} w_{ij} d_{j,t-1}$ as another predetermined regressor, which becomes much simpler than checking the self-consistent condition under duration dependence like Brock and Durlauf (2001b, Ch.4.2). One remark is that there is no endogeneity issue in this case because the random private utility $\varepsilon_{i,t}$ in the previous section is assumed to be exogenous and each individual makes decisions myopically. That is, the current decision is based on the current values of observable covariates $z_{i,t}$ only and thus no simultaneity issue arises. The following condition is on the interaction weight w_{ij}, where we let W be the $n \times n$ matrix whose (i, j)-th element is w_{ij}.

Assumption 6 (interaction weight matrix) *(i) The interaction weight matrix W is predetermined and correctly specified; (ii) each element of W is nonnegative and all the diagonal elements are zero; (iii) W is row normalized, i.e., $\sum_{j=1}^{n} w_{ij} = 1$ for all i; (iv) W is independent of v_i for all i, and W is not in the range space of $X = (x_{1,1}, \cdots, x_{n,T})'$.*

It is important to assume that the interaction weight matrix W is predetermined and independent of v_i. Pre-specifying W outside the model is an easy way to obtain a predetermined and exogenous W, which prevents any identification problem as pointed by Manski (1993)$-$the reflection problem. The independence assumption also guarantees the absence of endogeneity problems since all the diagonal elements of W are zero by construction and Assumption 3 holds. Keeping the interaction weight matrix W out of the covariate space prevents any possible multicollinearity problem between $x_{i,t}$ and $\sum_{j \neq i} w_{ij} d_{j,t-1}$ in the index structure. The row normalization condition is standard in spatial econometrics literature (e.g., Anselin and Bera, 1998), which prevents the weighted sum $\sum_{j \neq i} w_{ij} d_{j,t-1}$ from exploding under the in-filling asymptotics (i.e., increasing the number of observations within a fixed boundary). It controls the degree of cross sectional dependence so that the standard M-estimator has proper asymptotic properties (e.g., Kelejian and Prucha, 1999; Lee, 2004).

4 Estimation via EM Algorithm

4.1 Log-likelihood function

We let $c_{i,t}$ be the binary censoring indicator equal to one if the duration of i is censored in the t-th interval $[t-1,t)$. Apparently, $c_{i,t}=1$ implies $c_{i,t+1}=1$ for all i and the censoring variable C_i corresponds to the smallest t that gives $c_{i,t}=1$.

We consider the individual i, who drops out of the sample in the T_i-th interval $[T_i-1,T_i)$ either by exiting the initial state $(c_{i,T_i}=0)$ or by censoring $(c_{i,T_i}=1)$. If the panel is balanced, $T_i=T$ for all i. If we ignore conditioning on the initial state, the conditional log-likelihood function on the unobserved v is then given by

$$
\begin{aligned}
\log L\left(\gamma,\theta|v\right) = & \sum_{i=1}^{n}\delta_i\log\left(1-\exp\left(-\exp\left(\gamma_{T_i}\right)\phi\left(x'_{i,T_i}\beta+\alpha Wd_{T_i-1}\right)v_i\right)\right) \quad\quad (15)\\
& -\sum_{i=1}^{n}\sum_{s=1}^{T_i-1}\exp\left(\gamma_s\right)\phi\left(x'_{i,s}\beta+\alpha Wd_{s-1}\right)v_i
\end{aligned}
$$

from (14) similarly as Prentice and Gloeckler (1978) and Meyer (1990, 1995), where $\delta_i=1-c_{i,T_i}$, $\gamma=(\gamma_1,\gamma_2,\cdots,\gamma_T)'$ and $\theta=(\beta',\alpha)'$. We also denote $Wd_{t-1}=\sum_{j\neq i}w_{ij}d_{j,t-1}$ with $d_{t-1}=(d_{1,t-1},\cdots,d_{n,t-1})'$ for $w_{ii}=0$. Integrating (15) over the distribution of v yields the unconditional log-likelihood given by

$$
\log L\left(\gamma,\theta\right)=\sum_{j=1}^{m}p_j\log L\left(\gamma,\theta|v=q_j\right). \quad\quad (16)
$$

Note that, however, the ML estimation is not appropriate on the mixture model (16) since the parameters of the heterogeneity distribution are not guaranteed to lie on the interior of a compact set (e.g., Heckman and Singer, 1984; Lancaster, 1990, Chapter 8.4; Lee, 2000). Moreover, many studies report that the ML estimation of mixture models has convergence problem when the models have both the piecewise constant baseline hazard and the finite mixture unobserved heterogeneity.

For estimation, we instead use the EM (Expectation-Maximization) algorithm (e.g., Dempster, Laird and Rubin, 1977), which was originally invented to deal with inference in models

imposing missing data. In our case, the unobserved heterogeneity v is essentially a problem of missing data. To fix this idea, we consider an alternative expression of the conditional log-likelihood function (15) for individual i as

$$\log L_i \left(\gamma, \theta | v_i = q_j \right) = \sum_{j=1}^{m} \eta_{ij} \log L_i^* \left(\gamma, \theta, q_j \right), \tag{17}$$

where $\eta_{ij} = 1\{v_i = q_j\}$ for all $i = 1, \cdots, n$ and $j = 1, \cdots, m$ and

$$\begin{aligned}
\log L_i^* \left(\gamma, \theta, q_j \right) &= \delta_i \log \left(1 - \exp \left(- \exp \left(\gamma_{T_i} \right) \phi \left(x'_{i,T_i} \beta + \alpha W d_{T_i - 1} \right) q_j \right) \right) \\
&\quad - \sum_{s=1}^{T_i - 1} \exp \left(\gamma_s \right) \phi \left(x'_{i,s} \beta + \alpha W d_{s-1} \right) q_j.
\end{aligned} \tag{18}$$

Note that η_{ij} is unobservable and it can be viewed as missing data satisfying $\log f \left(v_i \right) = \sum_{j=1}^{m} \eta_{ij} \log p_j$ for the prior probabilities p_j $(j = 1, 2, \cdots, m)$ from (13). Hence the unconditional joint log-likelihood function of both the observed and the unobserved items is defined as

$$\log L \left(\Theta \right) = \sum_{i=1}^{n} \left\{ \log f \left(v_i \right) + \log L_i \left(\gamma, \theta | v_i \right) \right\} = \sum_{i=1}^{n} \sum_{j=1}^{m} \eta_{ij} \left\{ \log p_j + \log L_i^* \left(\gamma, \theta, q_j \right) \right\}, \tag{19}$$

where $\Theta = \left(\gamma', \theta', q_1, \cdots, q_m, p_1, \cdots, p_{m-1} \right)'$ is the complete parameter vector, provided that the known distribution function $G(\cdot)$ is free of additional unknown parameters. Note that p_m is automatically determined from the restriction of probabilities, $p_m = 1 - \sum_{j=1}^{m-1} p_j$.

4.2 EM algorithm

The EM algorithm proceeds in two steps. The E-step calculates the conditional expectation of η_{ij} given observed data δ_i and $Z_i = \{ z_{i,s} = (x'_{i,s}, W d_{s-1})' : 0 \leq s \leq T_i \}$ and given the likelihood L_i^* evaluated at the current parameter estimates. It can be shown that the posterior probability of $v_i = q_j$ is derived as

$$\mathbb{E} \left(\eta_{ij} | \delta_i, Z_i \right) = \frac{p_j L_i^* \left(\gamma, \theta, q_j \right)}{\sum_{\ell=1}^{m} p_\ell L_i^* \left(\gamma, \theta, q_\ell \right)} \equiv \pi_{ij} \quad \text{for all } i = 1, \cdots, n \text{ and } j = 1, \cdots, m. \tag{20}$$

If we let $\widehat{\pi}_{ij}$ denote π_{ij} evaluated at the current parameter estimates, substituting $\widehat{\pi}_{ij}$ for η_{ij} in (19) gives the outcome of the E-step:

$$Q\left(\Theta\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \widehat{\pi}_{ij} \log p_j + \sum_{i=1}^{n}\sum_{j=1}^{m} \widehat{\pi}_{ij} \log L_i^*\left(\gamma, \theta, q_j\right). \tag{21}$$

Then, the M-step consists of maximizing $Q\left(\Theta\right)$ with respect to Θ, which only requires numerical maximization of $\log L_i^*\left(\gamma, \theta, q_j\right)$. Maximization with respect to p_j's has an explicit solution as $\widehat{p}_j = n^{-1}\sum_{i=1}^{n} \widehat{\pi}_{ij}$ from a general result in finite mixture models (e.g., Everitt and Hand, 1981). We iterate the entire E and M-steps until the estimates converge.

The initial values for the EM algorithm can be chosen from the ML estimates of the duration model without unobserved heterogeneity $\lambda_i\left(t\right) = \exp(\gamma_t)\phi(x'_{i,t}\beta + \alpha W d_{t-1})$. We then start the EM algorithm on the duration model with unobserved heterogeneity $\lambda_i\left(t|u_i\right) = \exp(\gamma_t + u_i)\phi(x'_{i,t}\beta + \alpha W d_{t-1})$, where $v_i = \exp(u_i)$, using the first step ML estimates $(\gamma_1^0, \cdots, \gamma_T^0, \beta^0, \alpha^0)$ and arbitrary (q^0, p^0) as the initial values. Note that the reparametrization $v_i = \exp(u_i)$ is convenient in solving the maximization problem since we do not need to impose restrictions on the sign of unobserved heterogeneity (i.e., $0 < v_i < \infty$ holds for any u_i). For the distribution of v_i, we start with two points of support (q_1, q_2) and keep adding more points of support as long as all the estimates $\widehat{q}_j$ are distinct. Leaving the mean of v unrestricted, we omit the first term of baseline hazard (i.e., $\gamma_1 = 0$) so that the parameter to estimate is $\Theta = (\gamma_2, \cdots, \gamma_T, \beta', \alpha, q_1, \cdots, q_m, p_1, \cdots, p_{m-1})'$. The Simulated Annealing (e.g., Goffe et al., 1994) grid search is helpful to find or confirm the global maximum.

The EM algorithm is known to be robust to the choice of initial values and practically guarantees convergence to at least a local maximum. One of its disadvantages is that it dose not provide standard errors as an immediate by-product unlike the Newton-Raphson type methods. Louis (1982), Meng and Rubin (1991), Guo and Rodriguez (1992) and Oakes (1999) propose a way to find a positive-definite observed information matrix within the EM algorithm framework, from which we can obtain the asymptotic variance matrix. More precisely, under the regularity conditions, Louis (1982) shows that the observed information matrix $\mathcal{I}\left(\Theta\right)$ can be obtained as $\mathcal{I}\left(\Theta\right) = \mathbb{E}_v\left(\frac{\partial^2}{\partial\Theta\partial\Theta'}\log L\left(\Theta\right)\right) - \mathbb{V}_v\left(\frac{\partial}{\partial\Theta}\log L\left(\Theta\right)\right)$, where $L\left(\Theta\right)$ is the complete data likelihood

13

function in (19). The expectation $\mathbb{E}_v(\cdot)$ and variance $\mathbb{V}_v(\cdot)$ are taken over the conditional distribution of v given the observed data $\{\delta_i, Z_i\}_{i=1}^n$. As noted in Guo and Rodriguez (1992), the first term of $\mathcal{I}(\Theta)$ can be interpreted as the conditional expectation of the observed information when v is observed, whereas the second term represents the missing information associated with the conditional distribution of v given the observed data. The explicit form of the observed information matrix $\mathcal{I}(\Theta)$ in our model is given in the Appendix.

5 The Passage of the Motorcycle-Helmet-Use Law

5.1 Data

Motorcycle-helmet-use (MHU) law The history of each state's coverage of the MHU law is obtained from the Insurance Institute for Highway Safety (www.iihs.org/laws/helmet_-history.html). In our model $d_{i,t} = 1$ means state i chooses the universal MHU law while $d_{i,t} = -1$ denotes all choices other than the universal law, which includes not only repealing the law but also passing the partial MHU law. Note that the partial MHU law applies only to young riders under a certain age (e.g. not older than 17) and adult riders either inexperienced (e.g. instruction permit holders) or without sufficient medical insurance. So the proportion of riders covered by the partial law is rather small. Under such definition, we have 15 states with no revision and 33 states with revisions. More precisely, we have 27 states with one revision, 4 states with 2 revisions, and other two states with 3 and 4 revisions, respectively, during the study period. Among these 42 revisions, 34 cases are from -1 to 1 (i.e., adopt the universal law) whereas 8 cases are from 1 to -1 (i.e., repeal or reduce the universal law). Even though we have the dates of law changes, we set the unit of time as a month because of the availability of covariates, so that the entire sample period is 383 months (from February 1975 to December 2006; $T = 383$) for 48 states excluding Hawaii and Alaska ($n = 48$).

Though the study period is long, number of passages is limited and the baseline hazard γ_t does not change frequently. To estimate the piecewise constant baseline hazard (12) more efficiently, we group the time periods into four such that they reflect the history of legislating activities regarding to the MHU law during the sample period as follows (source:

14

www.iihs.org/research/qanda/helmet_use.html):

- *From Feb. 1975 to Dec. 1978 ($1 \leq t \leq 47$)*: In 1966, the Highway Safety Act was introduced by the federal government, which required states to have the mandatory MHU law if they wanted to receive federal funds for highway maintenance and construction. 47 states had complied by 1975; but in 1975, Congress withdrew it and half of the states had repealed the law within three years.

- *From Jan. 1979 to Dec. 1991 ($48 \leq t \leq 203$)*: There were no special activities to remark.

- *From Jan. 1992 to Sep. 1995 ($204 \leq t \leq 248$)*: In the Intermodal Surface Transportation Efficiency Act of 1991, signed by President Bush in December 1991, Congress created incentives for states to enact helmet and safety belt use laws. States with both laws were eligible for special safety grants, but states that had not enacted them by October 1993 had up to 3 percent of their federal highway allotment redirected to highway safety programs.

- *From Oct. 1995 to Dec. 2006 ($249 \leq t \leq 383$)*: Four years after establishing the incentives, Congress again reversed itself. In the fall of 1995, Congress lifted federal sanctions against states without MHU laws, paving the way for state legislatures to repeal the MHU laws.

Figure 1 shows the status of choices at four selected time periods, $t = 1, 47, 248$ and 383. Shaded states have no requirement or maintain the partial MHU law (i.e., $d_{i,t} = -1$).

[Figure 1 is about here]

Note that we define two types of spells and failures: "type -1 spell" ("type 1 spell") denotes a spell in which individual i stays with $d_{i,t} = -1$ ($d_{i,t} = 1$) throughout. An individual starts "type -1 spell" if her choice changes from 1 to -1 (i.e., "($1 : -1$)-failure") and starts "type 1 spell" if her choice changes from -1 to 1 (i.e., "($-1 : 1$)-failure"). Each spell ends by either a new choice (i.e., a failure) or censoring then the other type of spell starts. Figure 2-(a) depicts the Kaplan-Meier estimator for survivor function on both types of failures where the vertical lines separate the time period as described above. In more details, Figure 2-(b) plots the two Kaplan-Meier estimators by the type of failures separately. We observe that all the failures in the first period are in "($1 : -1$)-failure". Big drop at the beginning of the third period accounts

for a revision from a partial to universal coverage in California. (The records show that only two states expanded the coverage during the third period: California expanded the law to universal coverage in January 1992 and Rhode Island expanded coverage to operators 20 years-old and younger in July 1992. In our definition of revision, however, only California's revision counts.) Finally most revisions in the last period are in "$(1 : -1)$-failure". It suggests that when the Federal government lifted the incentive to adopt the universal MHU law, states reacted quickly to repeal or reduce the law as shown in the first and the last period. States, however, have the incentive to adopt the universal law voluntarily as shown in the second period group where no incentive or requirement was given by the Federal government.

[Figure 2 is about here]

Social interactions   To analyze the social interaction between neighboring states, we consider the 48 contiguous states excluding Hawaii and Alaska, where the neighborhood is defined based on the geographical locations. More precisely, if two states are adjacent or share borders, then they are defined as neighbors: $w_{ij} = w_{ji} = 1$ if i and j are neighbors to each other, $w_{ij} = 0$ otherwise. We naturally assume that the neighborhood is time invariant in this case. We redefine w_{ij} by dividing it by the total number of borders shared (i.e., total number of neighbors) of state i, to make it as a well-defined weight as well as row-normalized (i.e., $\sum_{j=1}^{n} w_{ij} = 1$).

Considering geographical locations as a measure of the social-dependence measure is intuitive in this particular example, even after controlling for any geographical, meteorological and cultural similarities among the states. Motorcyclists frequently traveling between state borders would like to have a homogenous regulation between them. Insurance companies may present the data comparing fatality rates across state border to emphasize that introducing the universal MHU law reduces the probability of mortality. If we regard a state's decision on the coverage law as a conclusion of the residents' consensus, therefore, the geographical proximity would be a strong factor to measure the social distance in this decision making procedure.

Control variables   National Highway Traffic Safety Administration (NHTSA) maintains the Fatality Analysis Reporting System (FARS), which contains monthly fatality data from 1975

by the vehicle types. In this study, we count fatalities occurred by motorcycles only excluding mopeds, mini-bikes and motor scooters, and use the fatality rates per population. Since each state can only observe the fatality rate up to previous month at each time, we include the fatality rate lagged by one month. The number of registered motorcycles is provided by the Federal Highway Administration (FHWA). National Climate Data Center (NCDC) by National Oceanic and Atmospheric Administration (NOAA) provides the mean number of days in a month with precipitation 0.01 inch or higher, which could control for seasonality.

TABLE I: DEFINITIONS OF VARIABLES AND SUMMARY STATISTICS

Variable	Description	Mean	Std.Dev	Min	Max
LRoadway	Total public road and street mileage in million miles in log scale	-2.750	0.795	-5.271	-1.184
Elevdiff	Difference between highest and lowest elevation in ft divided by 100,000	0.525	0.042	0.003	0.148
LPrecip	Mean number of days in a month with precipitation .01 inches or higher divided by 10 then logarithms taken	-0.133	0.367	-1.984	0.514
Population	Total population divided by a million	4.967	5.209	0.334	34.55
Registered	Number of registered motorcycles divided by 100,000	1.011	1.074	0.012	7.457
FatalRate	Number of fatalities in the previous month divided by Population	0.232	0.430	0	2.995
NbhdAvg	Average decisions of neighbors in the previous month	0.333	0.662	-1	1

TABLE II: Correlation between Variables

	LRoadway	ElevDiff	LPrecip	Population	Registered	FatalRate	NbhdAvg
LRoadway	1						
ElevDiff	0.094	1					
LPrecip	-0.188	-0.455	1				
Population	0.499	0.038	-0.084	1			
Registered	0.493	0.084	-0.147	0.844	1		
FatalRate	-0.319	0.031	0.001	-0.311	-0.262	1	
NbhdAvg	-0.103	-0.219	0.193	0.202	0.033	-0.224	1

Tables I and II summarize the variables used in the estimation and their descriptive statistics. *LRoadway*, *ElevDiff* and *LPrecip* represent the possibility of accident occurrences; *ElevDiff* and *LPrecip* represent the driving condition in each state. States have more incentive to adopt the universal MHU law for higher values of *LRoadway*, *ElevDiff* or *LPrecip*, and we expect the coefficients for these three covariates to be positive. *Population* and *Registered* represent the pressure of two conflicting opinions. Although they are naturally highly correlated, their effects on the choices of states are in the opposite direction. That is, as population grows, more car drivers are expected and so is higher pressure to introduce or maintain the universal MHU law. On the other hand, as the number of registered motorcycles rises, motorcyclists' opinion becomes more substantive. With regard to the state's decisions on the MHU law, the coefficient of *Population* is thus expected to be positive while that of *Registered* be negative.

The covariates *FatalRate* and *NbhdAvg* are of the particular interest. The fatality rate is expected to have a high positive impact on the hazard rate for "$(-1:1)$-failure." That is, if a state's safety concern is sufficiently substantial, it tends to have the universal MHU law when the previous fatality rate is high. Also, the effect of *NbhdAvg* is anticipated to be highly positive when the social interactions matters in this decision making. Note that the positive coefficient of *NbhdAvg* means that as more neighboring states adopt the universal MHU law, the higher hazard rate for "$(-1:1)$-failure" (i.e., revision of choice from -1 to 1) the state would encounter if it does not currently enforce the universal MHU law. At the same time, if the universal law is currently effective in the state, the hazard rate for "$(1:-1)$-failure" (i.e., revision of choice from 1 to -1) decreases with the positive coefficient of *NbhdAvg*. In other words, the probability rate that the state would repeal or reduce the universal law becomes lower as more neighbors adopt the universal law.

5.2   Estimation results

For estimation, we assume the logistic specification for $G(\cdot)$ in (7). More precisely, conditional on $z_{i,t}$, we assume that $\varepsilon_i(t,d)$ follows the i.i.d. type I extreme value distribution with scale parameter $\sigma > 0$ so that $G(z) = \exp(z/\sigma)/[1 + \exp(z/\sigma)]$ for all $-\infty < z < \infty$. For the identification purpose, we assume the unit scale parameter ($\sigma = 1$). In addition, since the

decisions can be in both directions (i.e., 1 to -1 by repealing the law and -1 to 1 by adopting the law), we specify the duration model as $\lambda_i(t|v_i) = \exp(\gamma_t)\phi_{d_{i,t-1}}(x'_{i,t}\beta + \alpha\sum_{j\neq i} w_{ij}d_{j,t-1})v_i$, where $\phi_{d_{i,t-1}}(z) = G(2z)$ if $d_{i,t-1} = -1$ and $1 - G(2z)$ if $d_{i,t-1} = 1$. Then the analysis becomes similar to a multi-spell duration analysis by letting the spell ends if any revision of the law is effective; the identification and estimation strategies in the previous sections can be readily extended. In this case, more precisely, the log-likelihood function is given by averages over all spells $k = 1, \cdots, K$ as

$$
\begin{aligned}
\log L\left(\gamma, \theta | v\right) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (22) \\
= \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{\substack{d_{i,T_i-1} \\ \in\{-1,1\}}} \delta_i(k)\log\left(1 - \exp\left(-\exp(\gamma_{\overline{T}_i(k)})\phi_{d_{i,\overline{T}_i(k)-1}}\left(x'_{i,\overline{T}_i(k)}\beta + \alpha W d_{\overline{T}_i(k)-1}\right)v_i\right)\right) \\
- \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{s=\underline{T}_i(k)}^{\overline{T}_i(k)-1}\sum_{\substack{d_{i,s-1} \\ \in\{-1,1\}}} \exp\left(\gamma_s\right)\phi_{d_{i,s-1}}\left(x'_{i,s}\beta + \alpha W d_{s-1}\right)v_i,
\end{aligned}
$$

where the k-th spell of individual i is from $\underline{T}_i(k)$ to $\overline{T}_i(k)$ and $\delta_i(k)$ is the censoring indicator of the spell k. In this context, we could understand the model as an alternating state model, where each spell does not rely on the history of events occurred before. Therefore, the interpretation of the sign of $(\alpha, \beta')'$ should be such that the positive values accelerate "$(-1 : 1)$-failure" that corresponds to the higher probability rate of switching from -1 to 1. Negative values, on the other hand, accelerate "$(1 : -1)$-failure" and decelerates the "$(-1 : 1)$-failure" at the same time. It thereby represents the lower probability rate of switching from -1 to 1.

Table III summarizes the estimation results. The first two columns show the ML estimates obtained without unobserved heterogeneity. When *NbhdAvg* is omitted, all estimates except for *Elevdiff* are highly significant but the signs for *LRoadway* and *FatalRate* are in the opposite direction to what were expected. Signs of all the estimates except for *LPrecip* and *FatalRate* are as expected when the social interaction *NbhdAvg* is taken into account. Note that, however, the estimates for *LPrecip* and (especially) *FatalRate* are not significant at the 5% level, while the effects from social interactions is highly significant.

19

TABLE III: ESTIMATION RESULTS

	No Unobs. Hetero. I	No Unobs. Hetero. II	With Unobs. Hetero.
LRoadway	-0.650**	0.354**	0.346**
	(0.059)	(0.073)	(0.057)
Elevdiff	3.549	3.243	1.759
	(2.170)	(2.134)	(2.099)
LPrecip	1.043**	-0.023	-0.599*
	(0.264)	(0.248)	(0.329)
Population	0.400**	0.255**	0.127**
	(0.015)	(0.067)	(0.022)
Registered	-1.579**	-1.519**	-0.705**
	(0.070)	(0.237)	(0.103)
FatalRate	-0.917**	-0.051	0.227
	(0.230)	(0.205)	(0.341)
NbhdAvg		4.124**	1.106**
		(0.327)	(0.232)
Base. Hazard: γ_2	-6.131**	-5.369**	-1.569**
	(0.016)	(0.013)	(0.358)
γ_3	-6.158**	-5.580**	-2.854**
	(0.248)	(0.488)	(0.563)
γ_4	-6.106**	-5.605**	-1.935**
	(0.010)	(0.023)	(0.352)
Unobs. Hetero.: q_1			0.036**
			(0.006)
q_2			0.000**
			(0.000)
p_1			0.688**
			(0.109)
Log-Likelihood	-767.588	-579.465	-238.370

Note: ML estimation is used for the case without unobserved heterogeneity, whereas EM method is used for the case with unobserved heterogeneity. Numbers in parentheses denote standard errors. * and ** represent significance at 10% and 5%, respectively.

The last column of the table displays the estimates of the MPH model assuming two levels of heterogeneity (q_1, q_2) using the EM algorithm with the simulated annealing to find the global maximum. The values on the second column are used for the initial values. We set $\gamma_1 = 0$, whereas the values of q_1 and q_2, and accordingly the mean of v, are left unrestricted. Signs of all

the estimates are as expected except for *LPrecip* though it is not significantly different from zero at the 5% level. For other covariates, *LRoadway* is the only significant variable that affects the decision of the MHU law. It suggests that the driving condition is not an important factor when the policy makers decide the level of the MHU law. However, as the total length of road grows, the state tends to have a stricter MHU law. Both parameters for covariates *Population* and *Registered*, which represent the pressure of two conflicting opinions, are significantly different from zero at the 5% significance level and are signed as expected.

Note that the sign of the estimates for *FatalRate* is as expected in the full model although it is not significant at the 5% level. However, the statistical insignificance does not mean that the MHU law is ineffective on the fatality rate. The estimation result in Table III only suggests that the previous fatality rate is not relevant when a state considers revision of the current MHU law. On the other hand, the estimates for *NbhdAvg* is significantly positive at the 5% level, which suggests that the policy makers in each state tend to make a parallel decision with neighboring states. It explains a behavioral aspect of the states' legislative decision making procedure (i.e., social interactions) and empirically shows how the proximity between states affects their decision making.

To show the magnitude of effect on hazard rate by each covariate, Table IV reports the averaged elasticities of hazard rates for each covariates. For log-transformed covariates (*LRoadway* and *LPrecip*), the values are obtained with respect to the percentage change of the original levels. For *NbhdAvg*, the values are obtained by the change in one of the neighbors' choices from -1 to 1 (e.g., Halvorsen and Palmquist, 1980). Note that elasticities do not depend on the baseline hazard by virtue of the proportional hazard specification. A 10% increase in total length of public roadway (*Roadway*) leads to 4.01% increase and 2.91% decrease in the hazard rate for "$(-1 : 1)$-failure" (i.e., likelihood to introduce the law) and "$(1 : -1)$-failure" (i.e., likelihood to repeal or reduce the law), respectively. Same movements are found for *Precip*, *Population* and *FatalRate*. The percentage change in the elevation difference (*Elevdiff*) would not induce much of percentage changes in hazard rates. On the other hand, 10% increase in the number of registered motorcycles (*Registered*) results in 9.12% decrease in hazard rate for "$(-1 : 1)$-failure" whereas 5.13% increase in that for "$(1 : -1)$-failure." Finally, when a neigh-

bor changes its choice from -1 to 1, 169.9% increase in hazard rate for "$(-1:1)$-failure" and 37.9% decrease in hazard rate for "$(1:-1)$-failure" are obtained, respectively. It shows that, as the elasticity by a neighbor's choice turns out substantially high, each state indeed reacts to the neighbors' decisions very sensitively. However, the degree of sensitivity is not symmetric: pressure from social interactions with neighboring states is higher, on the average, toward the direction to adopting the universal law than the other way around.

TABLE IV: ELASTICITIES OF HAZARD RATES

	Elasticity for "$(-1:1)$-failure"	Elasticity for "$(1:-1)$-failure"
Roadway	0.401	-0.291
Elevdiff	0.103	-0.082
Precip	-0.694	0.504
Population	0.706	-0.555
Registered	-0.912	0.513
FatalRate	0.312	-1.046
NbhdAvg	1.699	-0.379

Note: Letting $\lambda_i[d] = \exp(\gamma_t)\phi_d(x'_{i,t}\beta + \alpha\sum_{j\neq i} w_{ij}d_{j,t-1})v_i$ for $d = -1, 1$, the table represents averaged estimates of the elasticities $\partial \log\widehat{\lambda}_i[d]/\partial \log x_{i,t}$ over the observations. Elasticities for *Roadway* and *Precip* are calculated with respect to the change of *Roadway* and *Precip* before taking logs. Elasticities for *NbhdAvg* is obtained as (e.g., for the case of "(-1:1)-failure") $(\widehat{\lambda}_i^c[-1] - \widehat{\lambda}_i[-1])/\widehat{\lambda}_i[-1]$, where $\widehat{\lambda}_i^c[-1])$ is the counterfactual hazard rate when <u>one</u> of the neighbors changes from -1 to 1, whereas $\widehat{\lambda}_i[-1]$ is the original hazard rate estimate.

Finally, the estimation result of the baseline hazard in the full model with unobserved heterogeneity (the last column of Table III) shows that the piecewise baseline hazard estimate $(\exp(\widehat{\gamma}_1), \cdots, \exp(\widehat{\gamma}_4)) = (1.000, 0.208, 0.058, 0.144)$ are all highly significant. It well demonstrates the overall behavior of the states, which corresponds to the common trend of the historic changes of the federal regulations as summarized in the previous subsection. Note that the baseline hazard during the third period is substantially low, which shows that the incentives for states to enact the MHU law turned out to be not so effective.

For the levels of heterogeneity, the higher level $(\widehat{q}_1)$, which corresponds to the states that have higher tendency to change the law, is much larger than the lower level $(\widehat{q}_2)$ and the overall probability assigned to the former is about 2.2 times higher than that associated with the latter. As the estimates converge, the posterior probabilities $(\widehat{\pi}_{i1}, \widehat{\pi}_{i2})$ for each state i converges to the extreme values either 0 or 1. Figure 3 shows that the posterior probabilities correctly describe each state's behavior of changing the MHU law in that all states who have revised their law at least once are assigned to higher level of unobserved heterogeneity and vice versa. Note that in contrast to the standard duration data, in which a high-risk group is likely to have early events, a high-risk group in the longitudinal case appears to have frequent events, which can be described the faster rate of the Poisson process with higher level of v_i.

[Figure 3 is about here]

6   Concluding Remarks

The mandatory MHU law is a long-debated issue. One concerning safety supports the universal MHU law (e.g., Houston and Richardson, 2007; Muller, 2004; Weiss, 1992), which is mainly backed by automobile drivers (i.e., non-motorcyclists) and insurance companies. The opposite opinion adheres to the idea that wearing a helmet is a personal choice: Motorcyclists and liberalists insist that society's role is not to mandate personal safety but rather to provide the education and experience necessary to aid people in making these decisions for themselves. Moreover, even though it is a common wisdom that wearing helmets reduces the mortality in motorcycle accidents, medical evidences are still controversial (e.g., Cooter, et al., 1988; Goldstein, 1986; Huston and Sears, 1981; Krantz, 1985; Stolzenberg and D'Alessio, 2003).

Though we aware such a long debate on the mandatory MHU law, we do not attempt to answer to this question in this paper. The main point of this paper is to find the evidence that the coverage of the MHU law in a state is spatially dependent on neighboring states' coverage. We develop a model analyzing states' decision on the coverage of the MHU law. Reflecting the fact that each decision making can be only realized at certain times, a hazard model naturally follows. In particular, we introduce a grouped mixed proportional hazard (MPH) model with

social interactions, which serves as the econometric foundation of the states' choice model. Note that, however, this hazard model is different from the cross-sectional duration dependent models (e.g., Sirakaya, 2006) since the social interaction is based on others' discrete choice variables instead of the durations.

Based on the fact that not many revisions were made, we assume that the decisions are myopic. As a natural extension, a fully dynamic model incorporating a forward-looking behavior of agents is to be developed.

Appendix: Derivation of the observed information matrix

Louis (1982) shows that the observed information matrix $\mathcal{I}$ can be obtained as

$$\mathcal{I}(\Theta) = \mathbb{E}_v \left(\frac{\partial^2}{\partial\Theta\partial\Theta'} \log L(\Theta) \right) - \mathbb{V}_v \left(\frac{\partial}{\partial\Theta} \log L(\Theta) \right) \equiv \mathcal{I}_1(\Theta) - \mathcal{I}_2(\Theta).$$

We investigate these two terms separately. The multi-spell case in Section 5 can be handled similarly using the generalized log-likelihood (22).

Complete data information matrix $\mathcal{I}_1(\Theta)$ Under the regularity conditions, it can be obtained as $\mathcal{I}_1(\Theta) = -\partial^2 Q(\Theta)/\partial\Theta\partial\Theta'$ from the E-step. More precisely, from the formula (21) and for individual i,

$$Q_i(\Theta) = \sum_{j=1}^{m} \widehat{\pi}_{ij} \log p_j + \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \log\left(1 - \exp\left(-q_j\psi_{i,T_i}\right)\right) - \sum_{s=1}^{T_i-1} q_j\psi_{i,s} \right\},$$

where $\psi_{i,t} = \exp(\gamma_t)\phi(z'_{i,t}\theta)$. Since the maximization procedures of p and (γ, θ, q) can be separated, the Hessian matrix of $Q_i(\Theta)$ is block diagonal. In particular, for $t = 1, \cdots, T$ and $k = 1, \cdots, m$,

$$\frac{\partial}{\partial\gamma_t} Q_i(\Theta) = \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \frac{q_j\psi_{i,T_i} \exp\left(-q_j\psi_{i,T_i}\right)}{1 - \exp\left(-q_j\psi_{i,T_i}\right)} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j\psi_{i,s} 1\{t = s\} \right\},$$

$$\frac{\partial}{\partial\theta} Q_i(\Theta) = \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i \frac{q_j\psi_{i,T_i} \exp\left(-q_j\psi_{i,T_i}\right)}{1 - \exp\left(-q_j\psi_{i,T_i}\right)} \Psi^{(1)}_{i,T_i} - \sum_{s=1}^{T_i-1} q_j \Psi^{(1)}_{i,s} \right\},$$

$$\frac{\partial}{\partial q_k} Q_i(\Theta) = \widehat{\pi}_{ik} \left\{ \delta_i \frac{\psi_{i,T_i} \exp\left(-q_k\psi_{i,T_i}\right)}{1 - \exp\left(-q_k\psi_{i,T_i}\right)} - \sum_{s=1}^{T_i-1} \psi_{i,s} \right\},$$

$$\frac{\partial}{\partial p_k} Q_i(\Theta) = \frac{\widehat{\pi}_{ik}}{p_k} - \frac{\widehat{\pi}_{im}}{p_m} \quad (\text{for } k = 1, \cdots, m-1)$$

with $p_m = 1 - \sum_{j=1}^{m-1} p_j$, where $\Psi^{(1)}_{i,t} = \partial\psi_{i,t}/\partial\theta = \exp(\gamma_t)\phi^{(1)}(z'_{i,t}\theta)z_{i,t}$. We let

$$A_{ij} = \frac{\psi_{i,T_i} \exp\left(-q_j\psi_{i,T_i}\right)}{1 - \exp\left(-q_j\psi_{i,T_i}\right)}. \tag{A.1}$$

Then, for $t, r = 1, \cdots, T$ and $k, \ell = 1, \cdots, m$, we have

$$\frac{\partial^2 Q_i(\Theta)}{\partial\gamma_t\partial\gamma_r} = \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} \left(1 - q_j\psi_{i,T_i} - q_j A_{ij}\right) 1\{t = r = T_i\} - \sum_{s=1}^{T_i-1} q_j\psi_{i,s} 1\{t = r = s\} \right\},$$

$$\frac{\partial^2 Q_i(\Theta)}{\partial\theta\partial\gamma_t} = \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} \left(1 - q_j\psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}}\right) \Psi^{(1)}_{i,T_i} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j \Psi^{(1)}_{i,s} 1\{t = s\} \right\},$$

25

$$\frac{\partial^2 Q_i(\Theta)}{\partial\theta\partial\theta'} = \sum_{j=1}^{m} \widehat{\pi}_{ij} \left\{ \delta_i q_j A_{ij} \left[\left(1 - q_j \psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}} \right) \Psi_{i,T_i}^{(1)} \Psi_{i,T_i}^{(1)\prime} + \Psi_{i,T_i}^{(2)} \right] - \sum_{s=1}^{T_i-1} q_j \Psi_{i,s}^{(2)} \right\},$$

$$\frac{\partial^2 Q_i(\Theta)}{\partial\gamma_t\partial q_k} = \widehat{\pi}_{ik} \left\{ \delta_i A_{ij} \left(1 - q_j \psi_{i,T_i} - q_j A_{ij} \right) 1\{t = T_i\} - \sum_{s=1}^{T_i-1} \psi_{i,s} 1\{t = s\} \right\},$$

$$\frac{\partial^2 Q_i(\Theta)}{\partial\theta\partial q_k} = \widehat{\pi}_{ik} \left\{ \delta_i A_{ij} \left(1 - q_j \psi_{i,T_i} - q_j \frac{A_{ij}}{\psi_{i,T_i}} \right) \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i-1} \Psi_{i,s}^{(1)} \right\},$$

$$\frac{\partial^2 Q_i(\Theta)}{\partial q_k\partial q_\ell} = -\widehat{\pi}_{ik}\delta_i \left(\psi_{i,T_i} A_{ik} + A_{ik}^2 \right) 1\{k = \ell\},$$

$$\frac{\partial^2 Q_i(\Theta)}{\partial p_k\partial p_\ell} = -\frac{\widehat{\pi}_{ik}}{p_k^2} 1\{k = \ell\} + \frac{\widehat{\pi}_{im}}{p_m^2} \quad (\text{for } k, \ell = 1, \cdots, m-1),$$

where $\Psi_{i,t}^{(2)} = \partial^2 \psi_{i,t}/\partial\theta\partial\theta' = \exp(\gamma_t)\phi^{(2)}(z_{i,t}'\theta)z_{i,t}z_{i,t}'$. The complete data information matrix is then given by $\mathcal{I}_1(\Theta) = -\sum_{i=1}^{n} \partial^2 Q_i(\Theta)/\partial\Theta\partial\Theta'$, where $\partial^2 Q_i(\Theta)/\partial\Theta\partial\Theta'$ consists of the second derivatives obtained above conformably with the array of parameters

$$\Theta = \left(\gamma_1, \cdots, \gamma_T, \theta', q_1, \cdots, q_m, p_1, \cdots, p_{m-1} \right)'. \tag{A.2}$$

All other terms are zero since $\partial^2 Q_i(\Theta)/\partial\Theta\partial\Theta'$ is block diagonal.

Missing information matrix $\mathcal{I}_2(\Theta)$ From (19), we have

$$\log L_i(\Theta) = \sum_{j=1}^{m} \eta_{ij} \log p_j + \sum_{j=1}^{m} \eta_{ij} \left\{ \delta_i \log \left(1 - \exp\left(-\psi_{i,T_i} q_j \right) \right) - \sum_{s=1}^{T_i-1} \psi_{i,s} q_j \right\}$$

for individual i, which gives (for $t = 1, \cdots, T$ and $k = 1, \cdots, m$)

$$\frac{\partial}{\partial\gamma_t} \log L_i(\Theta) = \sum_{j=1}^{m} \eta_{ij} \left\{ \delta_i q_j A_{ij} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j \psi_{i,s} 1\{t = s\} \right\},$$

$$\frac{\partial}{\partial\theta} \log L_i(\Theta) = \sum_{j=1}^{m} \eta_{ij} \left\{ \delta_i q_j A_{ij} \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i-1} q_j \Psi_{i,s}^{(1)} \right\},$$

$$\frac{\partial}{\partial q_k} \log L_i(\Theta) = \eta_{ik} \left\{ \delta_i A_{ik} - \sum_{s=1}^{T_i-1} \psi_{i,s} \right\},$$

$$\frac{\partial}{\partial p_k} \log L_i(\Theta) = \frac{\eta_{ik}}{p_k} - \frac{\eta_{im}}{p_m} \quad (\text{for } k = 1, \cdots, m-1),$$

where A_{ij} is defined as (A.1). We let

$$
\begin{aligned}
B_{ij}^{\gamma_t} &= \delta_i q_j A_{ij} 1\{t = T_i\} - \sum_{s=1}^{T_i-1} q_j \psi_{i,s} 1\{t = s\}, \\
B_{ij}^{\theta} &= \delta_i q_j A_{ij} \Psi_{i,T_i}^{(1)} - \sum_{s=1}^{T_i-1} q_j \Psi_{i,s}^{(1)}, \\
B_{ij}^{q} &= \delta_i A_{ij} - \sum_{s=1}^{T_i-1} \psi_{i,s}.
\end{aligned}
$$

Then, since $Cov_v(\eta_{ik}, \eta_{i\ell}) = \pi_{ik}(1 - \pi_{ik}) 1\{k = \ell\}$ for $k, \ell = 1, \cdots, m$ from the definition of η_{ij}, where the covariance $Cov_v(\cdot)$ and the variance $\mathbb{V}_v(\cdot)$ are taken over the conditional distribution of v given the observed data $\{\delta_i, Z_i\}_{i=1}^n$, we derive that

$$
\begin{aligned}
\mathbb{V}_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}\right) &= \sum_{j=1}^m \pi_{ij}(1 - \pi_{ij}) B_{ij}^{\theta} B_{ij}^{\theta\prime}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial \gamma_s}\right) &= \sum_{j=1}^m \pi_{ij}(1 - \pi_{ij}) B_{ij}^{\gamma_t} B_{ij}^{\gamma_s} 1\{t = s\}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial q_k}, \frac{\partial \log L_i(\Theta)}{\partial q_\ell}\right) &= \pi_{ik}(1 - \pi_{ik}) \left(B_{ik}^q\right)^2 1\{k = \ell\}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial p_k}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{ik}(1 - \pi_{ik})}{p_k^2} 1\{k = \ell\} + \frac{\pi_{im}(1 - \pi_{im})}{p_m^2}
\end{aligned}
$$

for $t, r = 1, \cdots, T$ and $k, \ell = 1, \cdots, m$. Similarly,

$$
\begin{aligned}
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial \gamma_t}\right) &= \sum_{j=1}^m \pi_{ij}(1 - \pi_{ij}) B_{ij}^{\theta} B_{ij}^{\gamma_t}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial q_k}\right) &= \pi_{ik}(1 - \pi_{ik}) B_{ik}^{\theta} B_{ik}^q, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \theta}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{i\ell}(1 - \pi_{i\ell})}{p_\ell} B_{i\ell}^{\theta}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial q_k}\right) &= \pi_{ik}(1 - \pi_{ik}) B_{ik}^{\gamma_t} B_{ik}^q, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial \gamma_t}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{i\ell}(1 - \pi_{i\ell})}{p_\ell} B_{i\ell}^{\gamma_t}, \\
Cov_v\left(\frac{\partial \log L_i(\Theta)}{\partial q_k}, \frac{\partial \log L_i(\Theta)}{\partial p_\ell}\right) &= \frac{\pi_{ik}(1 - \pi_{ik})}{p_k} B_{ik}^q 1\{k = \ell\}
\end{aligned}
$$

for $t = 1, \cdots, T$, $k = 1, \cdots, m$ and $\ell = 1, \cdots, m - 1$. The missing information matrix is then given by $\mathcal{I}_2(\Theta) = \sum_{i=1}^n \mathbb{V}_v(\partial \log L_i(\Theta)/\partial \Theta)$, where $\mathbb{V}_v(\partial \log L_i(\Theta)/\partial \Theta)$ consists of covariance matrices obtained above conformably with the array of parameters Θ in (A.2). $\square$

References

Anselin, L. and A.K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics, in *Handbook of Applied Economics Statistics*, A. Ullah and D.E.A. Giles ed., New York: Marcel Dekker.

Brock, W. A., and S. N. Durlauf (2001a). Discrete choice with social interactions, *Review of Economic Studies*, 68, 235-260.

Brock, W. A., and S. N. Durlauf (2001b). Interactions-Based Models, in *Handbook of Econometrics*, vol. 5, ed. by J. J. Heckman, and E. Leamer, North-Holland: Amsterdam.

Carruthers, B., T. Guinnane, and Y. Lee (2010). Bringing Honest Capital to Poor Borrowers: The Passage of Uniform Small Loan Law, 1907-1930, *Journal of Interdisciplinary History*, forthcoming.

Cooper, R., and A. John (1988). Coordinating coordination failures in Keynesian models, *Quarterly Journal of Economics*, CIII, 441-464.

Cooter, R. D., A. J. McLean, D. J. David, and D. A. Simpson (1988). Helmet-Induced Skull Base Fracture in a Motorcyclist, *Lancet*, 1, 84-85.

Cox, D.R. (1972). Regression models and lifetime tables, *Journal of the Royal Statistical Society*, B 34, 187-220.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via EM algorithm, *Journal of the Royal Statistical Society*, Ser. B, 39, 1-38.

Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model, *Review of Economic Studies*, 49:3, 403-409.

Everitt, B.S. and D.J. Hand (1981). *Finite Mixture Distributions*, New York: Chapman and Hall.

Goffe, W.L., G.D. Ferrier and J. Rogers (1994). Global optimization of statistical functions with simulated annealing, *Journal of Econometrics*, 15:2, 265-287.

Goldstein, J. P. (1986). The Effect of Motorcycle Helmet Use on the Probability of Fatality and the Severity of Head and Neck Injuries, *Evaluation Review*, 10, 355-375.

Guo, G. and G. Rodriguez (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala, *Journal of the American Statistical Association*, 87, 969-976.

Halvorsen, R. and R. Palmquist (1980). The interpretation of dummy variables in semilogarithmic equations, *The American Economic Review*, 70, 474-475.

Heckman, L. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica*, 52:2, 271-320.

Houston, D. J., and L. E. Richardson (2007). Motorcycle Safety and the Repeal of Universal Helmet Laws, *American Journal of Public Health*, 97(11), 2063-2069.

Huston, R., and J. Sears (1981). Effect of Protective Helmet Mass on Head/Neck Dynamics, *Journal of Biomedical Engineering*, 103, 18-23.

Kelejian, H.H. and I.R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40, 509-533.

Kiefer, J. and J. Wolfowitz (1956). Consistency of the Maximum Likelihood Estimator in the presence of infinitely many incidental parameters, *The Annals of Mathematical Statistics*, 27:4, 887-906.

Kalbfleisch, J.D. and R.L. and Prentice (1973). Marginal likelihoods based on Cox's regression and life model, *Biometrika*, 60, 267-278.

Krantz, K. P. G. (1985). Head and Neck Injuries to Motorcycle and Moped Riders with Special Regard to the Effect of Protective Helmets, *Injury*, 16, 253-258.

Lancaster, T. (1979). Econometric methods for the duration of unemployment, *Econometrica*, 47, 939-956.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.

Lee, L.-F. (2000). A Numerically Stable Quadrature Procedure for the One-Factor Random-Component Discrete Choice Model, *Journal of Econometrics*, 95, 117-129.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica*, 72, 1899-1925.

Lee, Y. (2007). Grouped Mixed Proportional Hazard Models with Social Interactions, *mimeo*, University of Michigan.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society*, 44, 226-233.

Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, 60, 531-542.

Meng, X.L. and D.B. Rubin (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm, *Journal of the American Statistical Association*, 86, 899-909.

Meyer, B.D. (1990). Unemployment insurance and unemployment spells, *Econometrica*, 58:4, 757-782.

Meyer, B.D. (1995). Semiparametric Estimation of Hazard Models, *mimeo*, Northwestern University.

Muller, A. (2004). Florida's Motorcycle Helmet Law Repeal and Fatality Rates, *American Journal of Public Health*, 94(4), 556-558.

Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 61:2, 479-482.

Prentice, R. and L. Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data, *Biometrics*, 34, 57-67.

Sirakaya, S. (2006). Recidivism and social interactions, *Journal of the American Statistical Association*, 101, 863-877.

Stolzenberg, L., and S. J. D'Alessio (2003). "Born to Be Wild" the Effect of the Repeal of Florida's Mandatory Motorcycle Helmet-Use Law on Serious Injury and Fatality Rates, *Evaluation Review*, 27(2), 131-150.

Sueyoshi, G.T. (1995). A class of binary response models for grouped duration data, *Journal of Applied Econometrics*, 10, 411-431.

Trussell, J. and T. Richards (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure, in *Sociological Methodology 1985*, N. Tuma ed., San Francisco: Jossey-Bass, 242-276.

van den Berg, G.J. (2001). Duration Models: Specification, Identification, and Multiple Durations, in *Handbook of Econometrics*, vol. 5, J.J. Heckman and E. Leamer ed., North-Holland: Amsterdam.

Wallis, K. (1980). Econometric implications of the rational expectations hypothesis, *Econometrica*, 48, 49-73.

Weiss, A. A. (1992). The Effect of Helmet Use on the Severity of Head Injuries in Motorcycle Accidents, *Journal of the American Statistical Association*, 87, 48-56.
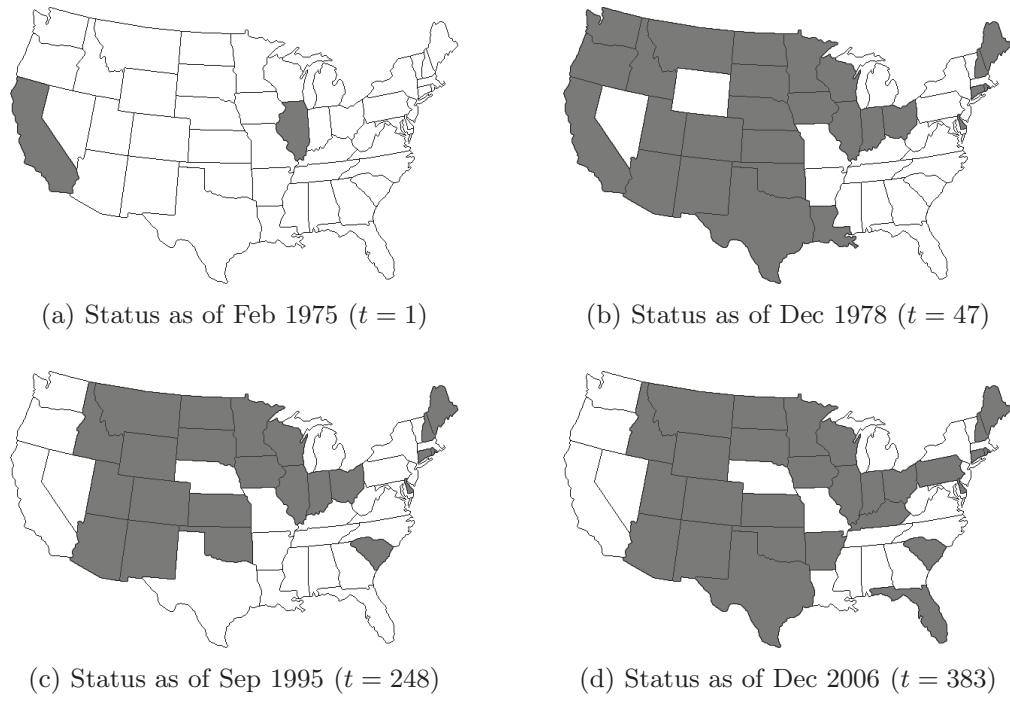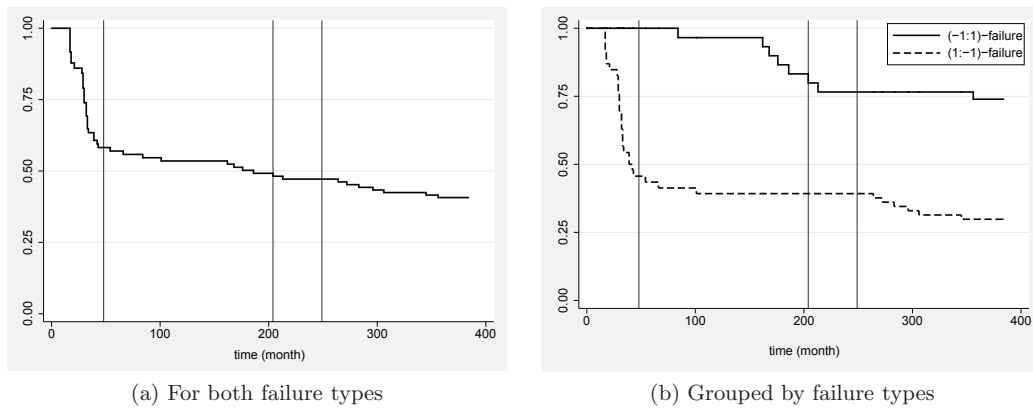
(a) Status as of Feb 1975 ($t = 1$)

(b) Status as of Dec 1978 ($t = 47$)

(c) Status as of Sep 1995 ($t = 248$)

(d) Status as of Dec 2006 ($t = 383$)

Figure 1. History of Helmet Use Law

(a) For both failure types

(b) Grouped by failure types

Figure 2. Kaplan-Meier Estimator for Survival Functions

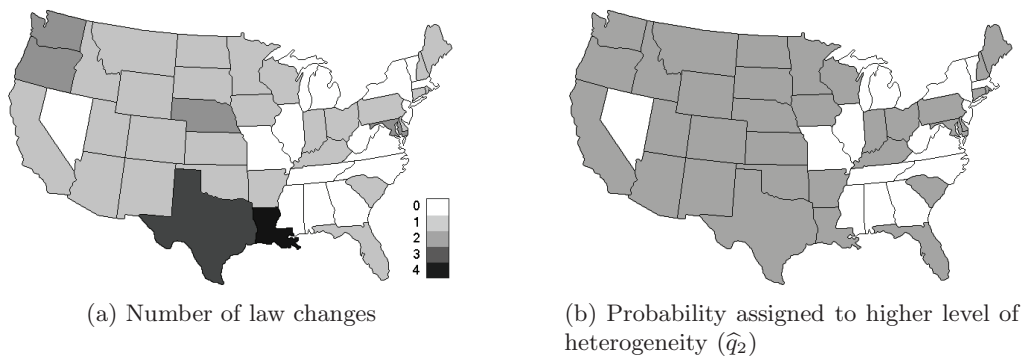(a) Number of law changes

(b) Probability assigned to higher level of heterogeneity $(\widehat{q}_2)$

Figure 3. Law Changes and Heterogeneity