Separate Appendix to: Semi-Nonparametric Competing Risks Analysis of Recidivism

Herman J. Bierens[a] * and Jose R. Carvalho[b]
[a] Department of Economics,Pennsylvania State University, University Park, PA 16802
[b] CAEN, Universidade Federal do Ceará, Fortaleza, Brazil

September 6, 2006

1 Identification

In this section we will re-derive the identification results of Heckman and Honore (1989) and Abbring and Van den Berg (2003)[1] for the common heterogeneity case, as follows.

1.1 Parametric identification

For $t \leq \overline{T}$, let the true conditional probability $P\left[T \leq t, D = 1, C = 0 | X\right]$ be

$$P\left[T \leq t, D = 1, C = 0 | X\right] \tag{1}$$
$$= \int_0^t h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}' X\right)\Lambda_{0,1}\left(\tau\right) + \exp\left(\beta_{0,2}' X\right)\Lambda_{0,2}\left(\tau\right)\right)\right)\right)$$

*The paper involved was presented by the first author at the Econometric Society European Meeting 2006 in Vienna. The helpful comments of Jaap Abbring are gratefully acknowledged.

[1] Abbring, J. H., and G. J. van den Berg (2003), "The Identifiability of the Mixed Proportional Hazards Competing Risks Model", *Journal of the Royal Statistical Society B*, 65, 701-710.

$$\times \exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(\tau)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(\tau)\right)\right)$$
$$\times \exp\left(\beta_{0,1}'X\right)\lambda_{0,1}(\tau)\,d\tau$$

Suppose there exist a density h on $[0,1]$, parameter vectors β_1, β_2 and hazard functions $\lambda_1(t)$ and $\lambda_2(t)$ with corresponding integrated hazards $\Lambda_1(t)$ and $\Lambda_2(t)$ such that for all $t \leq \overline{T}$,

$$\begin{aligned} &P[T \leq t, D = 1, C = 0|X] \\ &= \int_0^t h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(\tau)+\exp\left(\beta_2'X\right)\Lambda_2(\tau)\right)\right)\right) \\ &\times \exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(\tau)+\exp\left(\beta_2'X\right)\Lambda_2(\tau)\right)\right) \\ &\times \exp\left(\beta_1'X\right)\lambda_1(\tau)\,d\tau \end{aligned} \tag{2}$$

as well. Taking the derivative to t, it then follows that for all $t \leq \overline{T}$,

$$\begin{aligned} &h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right) \\ &\times \exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right) \\ &\times \exp\left(\beta_{0,1}'X\right)\lambda_{0,1}(t) \\ &= h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right)\right) \\ &\times \exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right) \\ &\times \exp\left(\beta_1'X\right)\lambda_1(t) \text{ a.s.} \end{aligned} \tag{3}$$

Similarly, if for all $t \leq \overline{T}$,

$$\begin{aligned} &P[T \leq t, D = 2, C = 0|X] \\ &= \int_0^t h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(\tau)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(\tau)\right)\right)\right) \\ &\times \exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(\tau)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(\tau)\right)\right) \\ &\times \exp\left(\beta_{0,2}'X\right)\lambda_{0,2}(\tau)\,d\tau \end{aligned}$$

is equal to

$$\begin{aligned} &P[T \leq t, D = 2, C = 0|X] \\ &= \int_0^t h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(\tau)+\exp\left(\beta_2'X\right)\Lambda_2(\tau)\right)\right)\right) \\ &\times \exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(\tau)+\exp\left(\beta_2'X\right)\Lambda_2(\tau)\right)\right) \\ &\times \exp\left(\beta_2'X\right)\lambda_2(\tau)\,d\tau \end{aligned}$$

then

$$
\begin{aligned}
& h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right) \\
& \times\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right) \\
& \times\exp\left(\beta_{0,2}'X\right)\lambda_{0,2}(t) \\
& = h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right)\right) \\
& \times\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right) \\
& \times\exp\left(\beta_2'X\right)\lambda_2(t) \ \text{ a.s.}
\end{aligned} \tag{4}
$$

Now suppose that $h_0(1) = h(1) = 1$, which corresponds to $E[V] = 1$. (See Assumption 2) Then, letting $t \downarrow 0$, it follows from (3) that

$$
\exp\left(\left(\beta_{0,1}-\beta_1\right)' X\right)\lim_{t\downarrow 0}\frac{\lambda_{0,1}(t)}{\lambda_1(t)} = 1 \text{ a.s.} \tag{5}
$$

If $\lambda_{0,1}(t)$ and $\lambda_1(t)$ are Weibull baseline hazards, including scale factors, i.e.,

$$
\lambda_{0,1}(t) = \alpha_{1,1}^*\alpha_{1,2}^*t^{\alpha_{1,2}^*-1},\ \ \lambda_1(t) = \alpha_{1,1}\alpha_{1,2}t^{\alpha_{1,2}-1}, \tag{6}
$$

where $\alpha_{1,1}^*$ and $\alpha_{1,1}$ are the scale factors involved, and all the parameters involved are positively valued, then

$$
\lim_{t\downarrow 0}\frac{\lambda_{0,1}(t)}{\lambda_1(t)} = \frac{\alpha_{1,1}^*}{\alpha_{1,1}}\times\frac{\alpha_{1,2}^*}{\alpha_{1,2}}\lim_{t\downarrow 0}t^{\alpha_{1,2}^*-\alpha_{1,2}} = \begin{cases} 0 & if \quad \alpha_{1,2}^* > \alpha_{1,2}, \\ \frac{\alpha_{1,1}^*}{\alpha_{1,1}} & if \quad \alpha_{1,2}^* = \alpha_{1,2}, \\ \infty & if \quad \alpha_{1,2}^* < \alpha_{1,2}. \end{cases}
$$

so that by (5), $\alpha_{1,2}^* = \alpha_{1,2}$ and

$$
X'\left(\beta_{0,1}-\beta_1\right) = \ln\left(\frac{\alpha_{1,1}}{\alpha_{1,1}^*}\right) \text{ a.s.} \tag{7}
$$

Because of the presence of the scale factors $\alpha_{1,1}^*$ and $\alpha_{1,1}$, we cannot allow a constant in X.

Next, suppose that the variance matrix $\Sigma_x = E\left[\left(X - E[X]\right)\left(X - E[X]\right)'\right]$ is non-singular (c.f. Assumption 3). Then it follows from (7) that $\Sigma_x\left(\beta_{0,1}-\beta_1\right) = 0$, hence $\beta_{0,1} = \beta_1$, and thus $\alpha_1 = \alpha_{0,1}$. Thus, in the case that the two baseline hazard functions are Weibull, Assumptions 2-3 guarantee the identification of the parameters.

In the case of non-Weibull baseline hazards we need more conditions. For example, suppose that

$$\lambda_1(t) = \frac{2\alpha_{1,1}t}{\alpha_{1,2}^2 + t^2},\ \lambda_{0,1}(t) = \frac{2\alpha_{1,1}^* t}{\left(\alpha_{1,2}^*\right)^2 + t^2}, \tag{8}$$

where again $\alpha_{1,1}^*$ and $\alpha_{1,1}$ are scale factors, and all the parameters are positively valued. These hazard functions are unimodal, with modes at $\alpha_{1,2} > 0$ and $\alpha_{1,2}^* > 0$, respectively. Then

$$\lim_{t \downarrow 0} \frac{\lambda_{0,1}(t)}{\lambda_1(t)} = \frac{\alpha_{1,1}^*}{\alpha_{1,1}} \times \frac{\alpha_{1,2}^2}{\left(\alpha_{1,2}^*\right)^2}, \tag{9}$$

hence (7) now becomes

$$(\beta_{0,1} - \beta_1)' X = \ln\left(\left(\alpha_{1,2}^*\right)^2 / \alpha_{1,2}^2\right) - \ln\left(\alpha_{1,1}/\alpha_{1,1}^*\right) \text{ a.s.} \tag{10}$$

Under Assumptions 2-3, (10) still implies that $\beta_{0,1} = \beta_1$ but now only that

$$\frac{\alpha_{1,2}^2}{\left(\alpha_{1,2}^*\right)^2} = \frac{\alpha_{1,1}}{\alpha_{1,1}^*}.$$

Thus, in the unimodal hazard case the Assumptions 2-3 do not guarantee identification of the parameters of the unimodal baseline hazard. It is easy to verify that the same problem occurs whenever

$$\lim_{t \downarrow 0} \lambda_{0,1}(t) / \lambda_1(t) \in (0, \infty) \backslash \{1\}$$

is possible. On the other hand, Assumptions 2-3 still guarantee that in the case $D = 1$, $\beta_{0,1} = \beta_1$, and similarly in the case $D = 2$ that $\beta_{0,2} = \beta_2$. Thus, (3) now reads,

$$\begin{aligned} & h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}' X\right) \Lambda_{0,1}(t) + \exp\left(\beta_{0,2}' X\right) \Lambda_{0,2}(t)\right)\right)\right) \\ & \times \exp\left(-\left(\exp\left(\beta_{0,1}' X\right) \Lambda_{0,1}(t) + \exp\left(\beta_{0,2}' X\right) \Lambda_{0,2}(t)\right)\right) \\ & \times \lambda_{0,1}(t) \\ & = h\left(\exp\left(-\left(\exp\left(\beta_1' X\right) \Lambda_1(t) + \exp\left(\beta_2' X\right) \Lambda_2(t)\right)\right)\right) \\ & \times \exp\left(-\left(\exp\left(\beta_{0,1}' X\right) \Lambda_1(t) + \exp\left(\beta_{0,2}' X\right) \Lambda_2(t)\right)\right) \\ & \times \lambda_1(t) \text{ a.s.} \end{aligned} \tag{11}$$

and (4) reads

$$\begin{aligned}
& h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right) \\
& \times\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right) \\
& \times\lambda_{0,2}(t) \\
& = h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right)\right) \\
& \times\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_1(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_2(t)\right)\right) \\
& \times\lambda_2(t)\ \text{ a.s.}
\end{aligned} \tag{12}$$

It follows from (11) and (12) that for all $t<\overline{T}$,

$$\frac{\lambda_2(t)}{\lambda_1(t)}=\frac{\lambda_{0,2}(t)}{\lambda_{0,1}(t)}. \tag{13}$$

To see what this result implies for the unimodal case, let similar to (8),

$$\lambda_2(t)=\frac{2\alpha_{2,1}t}{\alpha_{2,2}^2+t^2},\ \lambda_{0,2}(t)=\frac{2\alpha_{2,1}^*t}{\left(\alpha_{2,2}^*\right)^2+t^2}, \tag{14}$$

and assume that $\alpha_{1,2}^*\neq\alpha_{2,2}^*$, so that $\lambda_{0,1}(t)$ and $\lambda_{0,2}(t)$ are not proportional. Then it follows straightforwardly from (13), (8) and (14) that $\alpha_{1,2}=\alpha_{1,2}^*$, which implies that $\lambda_1(t)$ and $\lambda_{0,1}(t)$ are proportional, and therefore $\lambda_2(t)$ and $\lambda_{0,2}(t)$ are proportional as well, with common proportionality factor $c>0$, say:

$$\lambda_1(t)=c.\lambda_{0,1}(t),\ \lambda_2(t)=c.\lambda_{0,2}(t).$$

But then it follows from (11) and Assumption 2 that

$$\begin{aligned}
c &= \lim_{t\downarrow 0}\frac{h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right)}{h\left(\exp\left(-\left(\exp\left(\beta_1'X\right)\Lambda_1(t)+\exp\left(\beta_2'X\right)\Lambda_2(t)\right)\right)\right)} \\
&\quad\times\frac{\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)}{\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_1(t)+\exp\left(\beta_{0,2}'X\right)\Lambda_2(t)\right)\right)} \\
&= 1,
\end{aligned}$$

hence $\lambda_1(t)=\lambda_{0,1}(t),\ \lambda_2(t)=\lambda_{0,2}(t)$.

If $\alpha_{1,2}^*=\alpha_{2,2}^*$, which implies that for some constant $\kappa>0$, $\lambda_{0,2}(t)=\kappa\lambda_{0,1}(t)$, then (13) implies that $\lambda_2(t)=\kappa\lambda_1(t)$ as well, but not necessarily that (9) holds. Therefore, proportionality of $\lambda_{0,1}(t)$ and $\lambda_{0,2}(t)$ has to be excluded, at least for t close to zero:

Assumption A.1. *If the true baseline hazards* $\lambda_{0,1}(t)$ *and* $\lambda_{0,2}(t)$ *are non-Weibull, then they have to be non-proportional in the sense that there exists a small* $\varepsilon > 0$ *such that for any constant* $\kappa > 0$ *the set* $\{t \in (0, \varepsilon) : \lambda_{0,2}(t) = \kappa.\lambda_{0,1}(t)\}$ *has Lebesgue measure zero.*

In general (13) and Assumption A.1. are necessary but not sufficient conditions for (9), because we can always choose a hazards function $\lambda_1(t)$ such that $\lambda_2(t)$ defined by

$$\lambda_2(t) \equiv \left(\frac{\lambda_1(t)}{\lambda_{0,1}(t)}\right) \lambda_{0,2}(t). \tag{15}$$

is a valid hazard function. The reason that (9) holds for Weibull hazards and the unimodal hazards is that the four hazard functions involved have the same functional forms, which is such that (15) implies that $\lambda_1(t)$ and $\lambda_{0,1}(t)$ are proportional. Therefore, we need to require that

Assumption A.2. *If the baseline hazard functions in the competing risks model are non-Weibull then they belong to a class of parametric hazard functions* $\mathcal{L} = \{\lambda(t|\alpha), \alpha \in A\}$ *such that for any pair* $\lambda_{0,1}, \lambda_{0,2}$ *of non-proportional*[2] *hazard functions in* $\mathcal{L}$, (15) *can only hold for a pair* $\lambda_1, \lambda_2 \in \mathcal{L}$ *if and only if* $\lambda_1(t) \equiv c.\lambda_{0,1}(t)$ *for some constant* $c > 0$.

Summarizing, we have shown that

Theorem A.1. *If the baseline hazards are Weibull then under Assumptions 2-3 the parameters of the competing risks model are identified. If the baseline hazards are non-Weibull then parameter identification requires the additional conditions in Assumptions A.1 and A.2.*

1.2 Nonparametric identification

Under Assumptions 2-4 and A.1-2 it follows now from (3) that for $t \leq \overline{T}$,

$$\begin{aligned} & h_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t) + \exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right) \\ &= h\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}(t) + \exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}(t)\right)\right)\right) \end{aligned}$$

[2] As defined in Assumption A.1.

a.s. By a similar argument it can be shown that

$$\begin{aligned} & H_0\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}\left(\overline{T}\right)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}\left(\overline{T}\right)\right)\right)\right) \\ & = H\left(\exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}\left(\overline{T}\right)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}\left(\overline{T}\right)\right)\right)\right) \end{aligned}$$

a.s. Thus, denoting

$$\begin{aligned} U &= \exp\left(-\left(\exp\left(\beta_{0,1}'X\right)\Lambda_{0,1}\left(T\right)+\exp\left(\beta_{0,2}'X\right)\Lambda_{0,2}\left(T\right)\right)\right), \\ \underline{u} &= \inf_{P[U\leq u]>0} u, \end{aligned}$$

we have that

$$H\left(u\right)=H_0\left(u\right) \text{ a.e. on } (\underline{u},1]. \tag{16}$$

Therefore, at first sight it seems that in the case of right-censoring it may not be true that

$$H(u)=H_0(u) \text{ a.e. on } [0,1]. \tag{17}$$

This is not a problem if we adopt Assumption 1:

Theorem A.2. *Given Assumption 1, let* q_0 *be the smallest natural number for which there exists a* $\delta_0\in\mathbb{R}^{q_0}$ *such that* $h_0(u)=h_{q_0}(u|\delta_0)$ *a.e. Then* δ_0 *is unique: If for some* $\delta\in\mathbb{R}^{q_0}$, $h_{q_0}(u|\delta_0)=h_{q_0}(u|\delta)$ *a.e. on a set with positive Lebesgue measure, then* $\delta=\delta_0$. *Moreover, for any* $q>q_0$ *and* $\delta\in\mathbb{R}^q$ *such that* $h_0(u)=h_q(u|\delta)$ *a.e. on a set with positive Lebesgue measure, we have* $\delta'=(\delta_0',0')$.

Proof: This result follows straightforwardly from Theorem 4 in Bierens (2006 c).

However, Assumption 1 is not necessary for nonparametric identification, but is merely adopted because it allows for standard maximum likelihood inference. In the competing risks case with common unobserved heterogeneity (16) does imply (17), because for $t\geq 0$, $H(\exp(-t))=\int_0^\infty \exp(-t.v)dG(v)$ and $H_0(\exp(-t))=\int_0^\infty \exp(-t.v)dG_0(v)$ are Laplace transforms of the distributions $G(v)$ and $G_0(v)$, respectively.[3]

Lemma A.1. *Let* $H(u)=\int_0^\infty u^v dG(v)$ *and* $H_0(u)=\int_0^\infty u^v dG_0(v)$ *for* $u\in[0,1]$, *where* $G(v)$ *and* $G_0(v)$ *are distribution functions with non-negative*

[3]We are indebted to Jaap Abbring for suggesting this.

support. If $H(u) = H_0(u)$ a.e. on an arbitrary interval $(\underline{u}, \overline{u}) \subset (0,1)$ then $G(v) = G_0(v)$ a.e. on $[0,\infty)$, hence $H(u) = H_0(u)$ a.e. on $[0,1]$.

Proof: First observe that for $u \in (0,1)$ and non-negative integers m,

$$\sup_{v\geq 0} v^m u^{v-1} < \infty. \tag{18}$$

Take the derivative of $H(u)$ and $H_0(u)$ to $u \in (\underline{u}, \overline{u})$. Then by (18) and dominated convergence we may take the derivatives inside the integrals involved:

$$\int_0^\infty v u^{v-1} dG(v) = \int_0^\infty v u^{v-1} dG_0(v). \tag{19}$$

Multiply (19) by u, and then take the derivatives to $u \in (\underline{u}, \overline{u})$ again, which by (18) implies that

$$\int_0^\infty v^2 u^{v-1} dG(v) = \int_0^\infty v^2 u^{v-1} dG_0(v).$$

Repeating this procedure it follows by induction that

$$\int_0^\infty v^m u^v dG(v) = \int_0^\infty v^m u^v dG_0(v) \text{ for } m = 0,1,2,.... \tag{20}$$

hence

$$\int_0^\infty \sum_{m=0}^{k} \frac{(t.v)^m}{m!} u^v dG(v) = \int_0^\infty \sum_{m=0}^{k} \frac{(t.v)^m}{m!} u^v dG_0(v) \text{ for } k = 0,1,2,.... \tag{21}$$

Since

$$\begin{aligned} \sup_{k\geq 1} \left| \sum_{m=0}^{k} \frac{(t.v)^m}{m!} u^v \right| &\leq \sum_{m=0}^{\infty} \frac{(|t|\,.v)^m}{m!} \exp\left(-v.\ln(1/u)\right) \\ &= \exp\left(\left(|t| - \ln(1/u)\right).v\right) \\ &\leq 1 \text{ if } |t| < \ln(1/u) \end{aligned}$$

it follows from (21) and bounded convergence that

$$\int_0^\infty \exp\left(t.v\right) u^v dG(v) = \int_0^\infty \exp\left(t.v\right) u^v dG_0(v) \text{ for } |t| < \ln(1/u). \tag{22}$$

Now denote

$$F(x|u) = \frac{\int_0^x u^v dG(v)}{\int_0^\infty u^v dG(v)},\ F_0(x|u) = \frac{\int_0^x u^v dG_0(v)}{\int_0^\infty u^v dG_0(v)} \tag{23}$$

for $u \in (\underline{u}, \overline{u})$ and $x > 0$. Then it follows from (22) and (23) that

$$\int_0^\infty \exp(t.v)\, dF(v|u) = \int_0^\infty \exp(t.v)\, dF_0(v|u) \text{ for } |t| < \ln(1/u).$$

Hence it follows from the uniqueness of moment-generating functions that $F(x|u) = F_0(x|u)$ for $u \in (\underline{u}, \overline{u})$ and $x > 0$, and thus

$$\int_0^x u^v dG(v) = \int_0^x u^v dG_0(v). \tag{24}$$

Moreover, similar to (20) it follows from (24) that for $x > 0$, $m, k = 0, 1, 2, ...$ and $u \in (\underline{u}, \overline{u})$,

$$\int_0^x v^{m+k} u^v dG(v) = \int_0^x v^{m+k} u^v dG_0(v),$$

hence

$$\begin{aligned}
\int_0^x v^m dG(v) &= \int_0^x v^m \sum_{k=0}^\infty \frac{(v.\ln(1/u))^k}{k!} u^v dG(v) \\
&= \sum_{k=0}^\infty \frac{(\ln(1/u))^k}{k!} \int_0^x v^{m+k} u^v dG(v) \\
&= \sum_{k=0}^\infty \frac{(\ln(1/u))^k}{k!} \int_0^x v^{m+k} u^v dG_0(v) \\
&= \int_0^x v^m dG_0(v).
\end{aligned}$$

Thus, for $x > 0$ and $m = 0, 1, 2, ...,$

$$\int_0^\infty (v.I(v \leq x))^m\, dG(v) = \int_0^\infty (v.I(v \leq x))^m\, dG_0(v), \tag{25}$$

where $I(.)$ is the indicator function.

Now use the well-known fact that distributions of bounded random variables are equal if and only if all their moments are equal. Then, with V a random drawing from $G(v)$ and V_0 a random drawing from $G_0(v)$, it follows from (25) that for $x, y > 0$,

$$P\left[V.I(V \leq x) \leq y\right] = P\left[V_0.I(V_0 \leq x) \leq y\right].$$

This implies that for $0 < y < x$, $G(x) - G(y) = G_0(x) - G_0(y)$. Hence, letting $x \to \infty$, it follows that $G(y) = G_0(y)$ for $y > 0$ and thus by right-continuity of distribution functions,

$$G(v) = G_0(v) \text{ for } v \geq 0.$$

Q.E.D.

2 Empirical results

2.1 Initial estimation and test results

Table A.1: Initial ML results

Parameters	$F = 0$		$F = 1$	
$(i = F + 1)$	*Estimates*	*t-values*	*Estimates*	*t-values*
$\beta_{i,1}$ $(MALE)$	0.134998	3.093	0.277564	4.969
$\beta_{i,2}$ $(BLACK)$	0.118005	4.449	0.284026	8.666
$\beta_{i,3}$ $(RELEASE)$	−0.336527	−12.082	0.266712	5.327
$\beta_{i,4}$ (AGE)	−0.066320	−11.168	−0.082643	−12.087
$\beta_{i,5}$ $(SENT)$	−0.190192	−9.679	−0.245142	−8.646
$\alpha_{i,1}$	0.997661	4.545	0.349503	4.543
$\alpha_{i,2}$	0.841034	25.182	0.779186	22.096

$q = 6$ $N = 15434$ $L.L. = -17640.2$

Table A.2: Logit results for felony arrest, $F = 1$

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	-0.061848	-0.092903	0.031055	0.016157	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.142566	0.236108	-0.093542	0.067847	MALE
$\beta_{2,2} - \beta_{1,2}$	0.166021	0.185658	-0.019637	0.042100	BLACK
$\beta_{2,3} - \beta_{1,3}$	0.603239	0.692580	-0.089341	0.064426	RELEASE
$\beta_{2,4} - \beta_{1,4}$	-0.016323	-0.011251	-0.005072	0.007923	AGE
$\beta_{2,5} - \beta_{1,5}$	-0.054950	-0.040778	-0.014172	0.032622	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	-1.125284	-1.405182	0.279898	0.125351	1
$n = 9979$	$L.L. = -6512.0$	ICM test: 6.73			

Table A.3: ML results with state fixed effects

Parameters	$F = 0$		$F = 1$	
$(i = F + 1)$	*Estimates*	*t-values*	*Estimates*	*t-values*
$\beta_{i,1}$ (*MALE*)	0.269285	4.956	0.319976	4.880
$\beta_{i,2}$ (*BLACK*)	0.216934	4.610	0.494636	9.195
$\beta_{i,3}$ (*RELEASE*)	-0.251983	-5.219	0.116589	1.674
$\beta_{i,4}$ (*AGE*)	-0.073790	-7.361	-0.082138	-7.422
$\beta_{i,5}$ (*SENT*)	-0.166639	-5.621	-0.255569	-6.655
$\beta_{i,6}$ (*Florida*)	0.111509	1.843	0.562870	6.130
$\beta_{i,7}$ (*Illinois*)	0.198016	2.955	0.779022	9.097
$\beta_{i,8}$ (*Michigan*)	-0.867757	-13.151	1.026690	15.436
$\beta_{i,9}$ (*Minnesota*)	-0.980232	-12.657	1.305132	19.488
$\beta_{i,10}$ (*New Jersey*)	-0.093203	-1.449	0.860569	11.432
$\beta_{i,11}$ (*New York*)	-0.350683	-5.932	0.796000	11.042
$\beta_{i,12}$ (*Ohio*)	-1.014663	-11.392	0.110678	1.074
$\beta_{i,13}$ (*Oregon*)	0.100728	1.102	1.445009	14.759
$\alpha_{i,1}$	0.889700	1.705	0.152724	1.732
$\alpha_{i,2}$	0.813537	8.994	0.748248	8.217
$q = 6$ $N = 15434$	$L.L = -16846.0$			

Table A.3: Logit results for felony arrest, $F = 1$, with state fixed effects

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.065289	−0.111763	0.046474	0.017140	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.050691	0.102199	−0.051508	0.071883	MALE
$\beta_{2,2} - \beta_{1,2}$	0.277702	0.319848	−0.042146	0.047113	BLACK
$\beta_{2,3} - \beta_{1,3}$	0.368572	0.398610	−0.030038	0.077124	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.008348	−0.002415	−0.005933	0.008433	AGE
$\beta_{2,5} - \beta_{1,5}$	−0.088930	−0.097458	0.008528	0.036732	SENT
$\beta_{2,6} - \beta_{1,6}$	0.451361	0.523263	−0.071902	0.092165	Florida
$\beta_{2,7} - \beta_{1,7}$	0.581006	0.649275	−0.068269	0.084797	Illinois
$\beta_{2,8} - \beta_{1,8}$	1.894447	1.973462	−0.079015	0.087442	Michigan
$\beta_{2,9} - \beta_{1,9}$	2.285364	2.345029	−0.059665	0.092384	Minnesota
$\beta_{2,10} - \beta_{1,10}$	0.953772	1.022188	−0.068416	0.082689	New Jersey
$\beta_{2,11} - \beta_{1,11}$	1.146683	1.265816	−0.119133	0.083556	New York
$\beta_{2,12} - \beta_{1,12}$	1.125341	1.132969	−0.007628	0.089646	Ohio
$\beta_{2,13} - \beta_{1,13}$	1.344281	1.412890	−0.068609	0.088791	Oregon
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−1.845909	−2.122809	0.276900	0.144235	1
$n = 9979$	$L.L. = -5949.9$	ICM test: 46.55			

2.2 Results per state

Table A.4: Logit results for felony arrest, $F = 1$: California

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.045997	−0.082745	0.036748	0.057254	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.281053	0.170924	0.110129	0.194597	MALE
$\beta_{2,2} - \beta_{1,2}$	0.676377	0.654849	0.021528	0.157434	BLACK
$\beta_{2,3} - \beta_{1,3}$	*N.A.*	*N.A.*	*N.A*	*N.A*	RELEASE
$\beta_{2,4} - \beta_{1,4}$	0.035060	−0.002233	0.037293	0.030352	AGE
$\beta_{2,5} - \beta_{1,5}$	−0.053146	0.008334	−0.061480	0.170823	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.250169	0.120246	−0.370415	0.390216	1
$n = 817$	$LL. = -516.17$	ICM test: 1.45			

Table A.5: Logit results for felony arrest, $F=1$: Florida

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2}-\alpha_{1,2}$	0.079968	0.043099	0.036869	0.046477	$\ln(T)$
$\beta_{2,1}-\beta_{1,1}$	−0.108228	0.058971	−0.167199	0.218233	MALE
$\beta_{2,2}-\beta_{1,2}$	0.438159	0.489171	−0.051012	0.138242	BLACK
$\beta_{2,3}-\beta_{1,3}$	−0.125350	−0.063814	−0.061536	0.146032	RELEASE
$\beta_{2,4}-\beta_{1,4}$	−0.079152	−0.026927	−0.052225	0.024198	AGE
$\beta_{2,5}-\beta_{1,5}$	−0.281664	−0.228791	−0.052873	0.171028	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.117836	−0.908198	0.790362	0.350965	1
$n=1150$	$LL.=-643.37$	ICM test: 1.31			

Table A.6: Logit results for felony arrest, $F=1$: Illinois

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2}-\alpha_{1,2}$	−0.050694	−0.053959	0.003265	0.055447	$\ln(T)$
$\beta_{2,1}-\beta_{1,1}$	−0.377491	−0.319539	−0.057952	0.216534	MALE
$\beta_{2,2}-\beta_{1,2}$	0.033885	0.066266	−0.032381	0.142627	BLACK
$\beta_{2,3}-\beta_{1,3}$	−0.412788	−0.328727	−0.084061	0.234898	RELEASE
$\beta_{2,4}-\beta_{1,4}$	−0.003869	−0.010677	0.006808	0.026747	AGE
$\beta_{2,5}-\beta_{1,5}$	0.111481	0.109510	0.001971	0.094837	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.103931	−0.207860	0.103929	0.440794	1
$n=960$	$L.L.=-606.04$	ICM test: 0.93			

Table A.7: Logit results for felony arrest, $F=1$: Michigan

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2}-\alpha_{1,2}$	0.074797	0.051207	0.023590	0.064258	$\ln(T)$
$\beta_{2,1}-\beta_{1,1}$	0.045925	0.105395	−0.059470	0.239307	MALE
$\beta_{2,2}-\beta_{1,2}$	0.649750	0.634523	0.015227	0.149179	BLACK
$\beta_{2,3}-\beta_{1,3}$	0.596651	0.629761	−0.033110	0.311634	RELEASE
$\beta_{2,4}-\beta_{1,4}$	−0.028903	−0.041851	0.012948	0.030639	AGE
$\beta_{2,5}-\beta_{1,5}$	0.041330	0.039043	0.002287	0.094718	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.087496	−0.065577	−0.021919	0.528535	1
$n=843$	$L.L.=-541.17$	ICM test: 1.06			

Table A.8: Logit results for felony arrest, $F = 1$: Minnesota					
Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.157274	−0.731430	0.574156	0.181235	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	−0.049170	−0.433112	0.383942	0.448080	MALE
$\beta_{2,2} - \beta_{1,2}$	0.834078	0.840727	−0.006649	0.213076	BLACK
$\beta_{2,3} - \beta_{1,3}$	0.680047	0.461176	0.218871	0.262323	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.025759	−0.058130	0.032371	0.029311	AGE
$\beta_{2,5} - \beta_{1,5}$	−0.174169	−0.225437	0.051268	0.180944	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	0.201754	1.689942	−1.488188	0.581945	1
$n = 805$	$L.L. = -471.94$	ICM test: 1.06			

Table A.9: Logit results for felony arrest, $F = 1$: New Jersey					
Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.172330	−0.214978	0.042648	0.055662	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.767053	0.773807	−0.006754	0.251017	MALE
$\beta_{2,2} - \beta_{1,2}$	0.023782	0.102342	−0.078560	0.139927	BLACK
$\beta_{2,3} - \beta_{1,3}$	−0.226671	−0.147122	−0.079549	0.430067	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.069330	−0.086143	0.016813	0.028229	AGE
$\beta_{2,5} - \beta_{1,5}$	0.014186	0.033818	−0.019632	0.123160	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.411254	−0.444947	0.033693	0.593683	1
$n = 941$	$L.L. = -618.57$	ICM test: 1.54			

Table A.10: Logit results for felony arrest, $F = 1$: New York					
Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.003175	−0.047414	0.044239	0.056983	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.586124	0.673544	−0.087420	0.231027	MALE
$\beta_{2,2} - \beta_{1,2}$	0.278358	0.324822	−0.046464	0.138372	BLACK
$\beta_{2,3} - \beta_{1,3}$	−0.196703	−0.282309	0.085606	0.364785	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.044891	−0.066334	0.021443	0.028680	AGE
$\beta_{2,5} - \beta_{1,5}$	−0.104860	−0.102353	−0.002507	0.091210	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.157937	−0.049243	−0.108694	0.528748	1
$n = 948$	$L.L. = -642.32$	ICM test: 1.70			

Table A.11: Logit results for felony arrest, $F = 1$: Ohio

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	0.210079	0.173498	0.036581	0.057842	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	−0.097975	−0.113565	0.015590	0.216840	MALE
$\beta_{2,2} - \beta_{1,2}$	0.598512	0.585218	0.013294	0.160185	BLACK
$\beta_{2,3} - \beta_{1,3}$	−0.202685	−0.142297	−0.060388	0.244004	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.081738	−0.088072	0.006334	0.032754	AGE
$\beta_{2,5} - \beta_{1,5}$	−0.153745	−0.197976	0.044231	0.142069	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	0.750235	0.740523	0.009712	0.420133	1
$n = 736$	$L.L. = -485.77$	ICM test: 1.44			

Table A.12: Logit results for felony arrest, $F = 1$: Oregon

Parameters	*Original*	*Logit*	*Difference*	*Logit s.e.*	*Variables*
$\alpha_{2,2} - \alpha_{1,2}$	−0.157469	−0.219420	0.061951	0.056864	$\ln(T)$
$\beta_{2,1} - \beta_{1,1}$	0.420576	0.595878	−0.175302	0.277425	MALE
$\beta_{2,2} - \beta_{1,2}$	0.924260	0.960285	−0.036025	0.242230	BLACK
$\beta_{2,3} - \beta_{1,3}$	0.031228	0.257992	−0.226764	0.348356	RELEASE
$\beta_{2,4} - \beta_{1,4}$	−0.047424	−0.027770	−0.019654	0.026407	AGE
$\beta_{2,5} - \beta_{1,5}$	0.202042	0.157533	0.044509	0.175693	SENT
$\ln\left(\frac{\alpha_{2,1}\alpha_{2,2}}{\alpha_{1,1}\alpha_{1,2}}\right)$	−0.472758	−1.131476	0.658718	0.532203	1
$n = 784$	$L.L. = -523.19$	ICM test: 1.84			

2.3 The SNP densities

The following plots of the SNP densities $h_q\left(u|\widehat{\delta}\right)$ have the same scale, $[0,9]\times[0,1]$, in order to make them comparable. The flatter the density, the less dependent the misdemeanor and felony recidivism durations are, conditional on the covariates.

To explain the shape of these densities, recall that the true density $h(u)$ is

$$h(u) = \int_0^\infty v u^{v-1} dG(v),$$

where $G(v)$ is the distribution function of the common unobserved heterogeneity variable V. Also, recall that the identification condition $E[V] = 1$ corresponds to $h(1) = 1$, which has been imposed on $h_q\left(u|\widehat{\delta}\right)$ as well. Therefore, $h_q\left(1|\widehat{\delta}\right) = 1$. Moreover, $E[V] = 1$ and $P[V = 1] < 1$ imply that $P[V < 1] > 0$, which in its turn implies that $\lim_{u\downarrow 0} h(u) = \infty$. Although this limit cannot be attained by $h_q\left(u|\widehat{\delta}\right)$ for finite q, it explains the shape of $h_q\left(u|\widehat{\delta}\right)$ close to $u = 0$.

Moreover, it seems that for California, Illinois, Michigan and New York the density $h_q\left(u|\widehat{\delta}\right)$ is zero for a $u \in (0,1)$, whereas the true density $h(u)$ cannot be zero, because $h(u) = 0$ for some point $u \in (0,1)$ implies $P[V = 0] = 1$, which violates the condition $E[V] = 1$. However, in these cases the minimum value of $h_q\left(u|\widehat{\delta}\right)$ is very small but positive. For example, in the case of Michigan $u_0 = \arg\min_{0\leq u\leq 1} h_4\left(u|\widehat{\delta}\right) = 0.18$, with $h_4\left(u_0|\widehat{\delta}\right) = 0.000616$.

To explain this phenomenon, suppose that V takes only two values,

$$P[V = \lambda] = p,\ P[V = \mu] = 1 - p$$

where $0 < \lambda < 1 < \mu < \infty$. The condition $E[V] = 1$ implies

$$\mu = \frac{1 - \lambda p}{1 - p}$$

hence

$$h(u) = \lambda p.u^{\lambda-1} + (1 - \lambda p)u^{(1-\lambda)p/(1-p)} \tag{26}$$

The first-order condition for an extremum of $h(u)$ in u_0 is

$$0 = -\lambda\left(1-\lambda\right)p.u_0^{\lambda-2} + (1-\lambda)\frac{p(1-\lambda p)}{1-p}u_0^{(1-\lambda)p/(1-p)-1}$$

hence

$$u_0 = \left(\frac{\lambda\left(1-p\right)}{1-\lambda p}\right)^{(1-p)/(1-\lambda)} = \left(\left(\frac{1-p}{1-\lambda p}\right)^{1-p}\lambda^{1-p}\right)^{1/(1-\lambda)} \tag{27}$$

Substituting this expression in (26) yields

$$h\left(u_0\right) = \left(\frac{1-\lambda p}{1-p}\right)^{1-p}\lambda^{p} = \lambda.u_0^{\lambda-1} \tag{28}$$

To show that $h(u_0)$ can get close to zero for some u_0 bounded away from zero, let for a given constant $c \in (0,1)$,

$$\lambda = c^{1/(1-p)}.$$

Then

$$\begin{aligned} \lim_{p \uparrow 1} u_0 &= \lim_{p \uparrow 1} \left(\left(\frac{1-p}{1 - c^{1/(1-p)} p} \right)^{1-p} c \right)^{1/(1-c^{1/(1-p)})} \\ &= c. \lim_{p \uparrow 1} (1-p)^{1-p} = c \end{aligned}$$

and

$$\lim_{p \uparrow 1} h(u_0) = \lim_{\lambda \downarrow 0} \lambda / c^{1-\lambda} = 0$$

More generally, if $h_q\left(u|\widehat{\delta}\right)$ takes a minimum in u_0 and $h_q\left(u_0|\widehat{\delta}\right)$ is small then λ and p can be chosen such that the density (26) takes a minimum in u_0, with $h_q\left(u_0|\widehat{\delta}\right) = h(u_0)$.

Of course, this is not the complete story, because in all cases $h_q\left(u|\widehat{\delta}\right)$ has two extrema rather than only one in this example. However, it is too difficult to construct a distribution of V that can explain the second extremum as well.

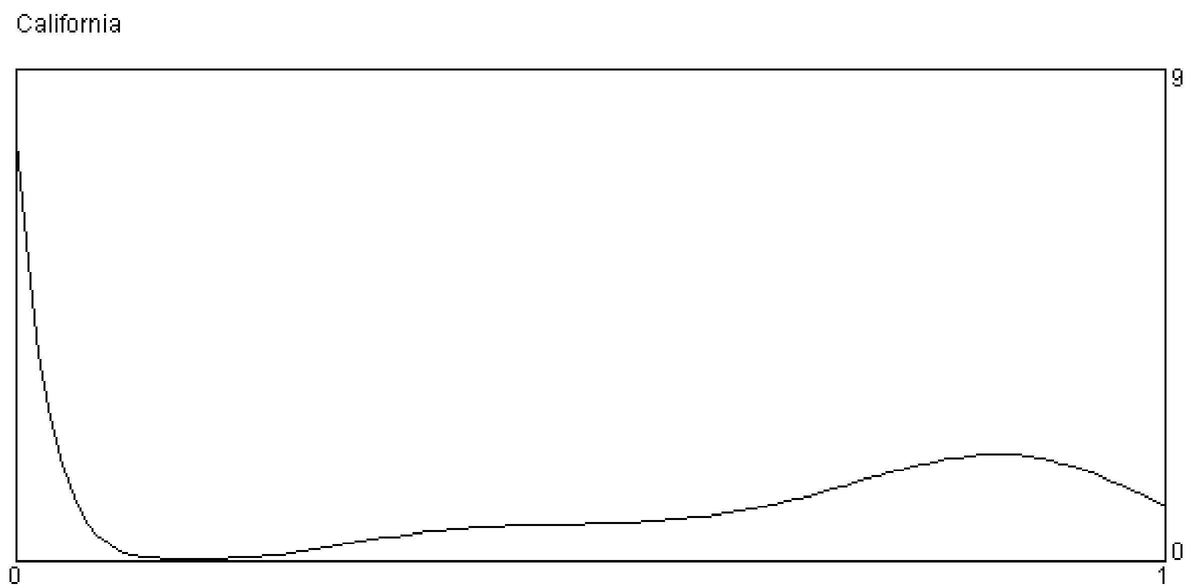

Figure 1: SNP density $h_6\left(u|\widehat{\delta}\right)$ for California

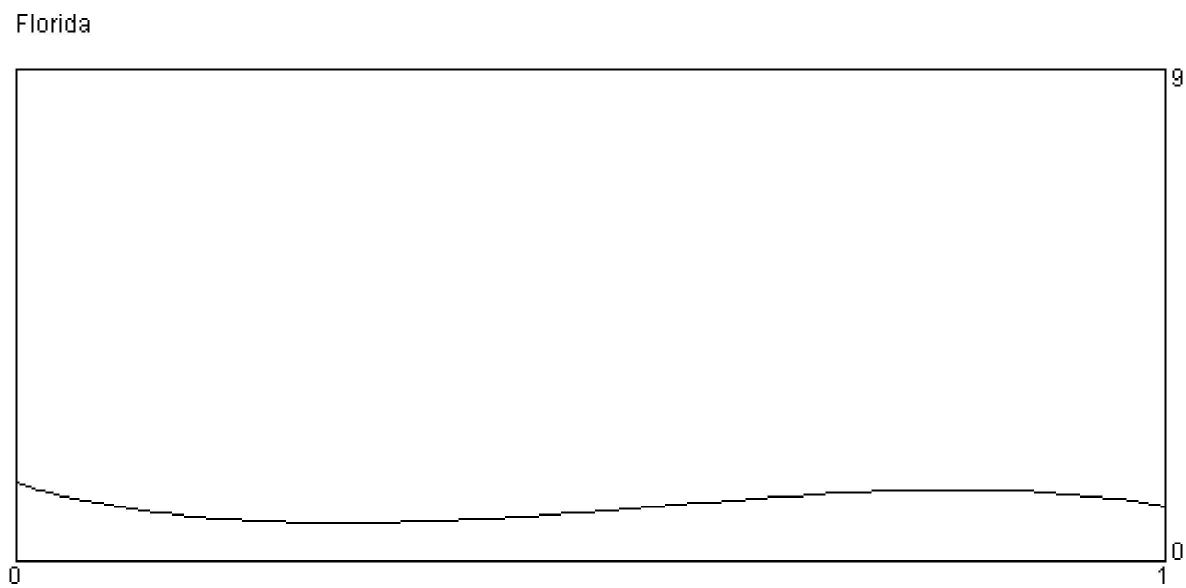

Figure 2: SNP density $h_3\left(u|\widehat{\delta}\right)$ for Florida

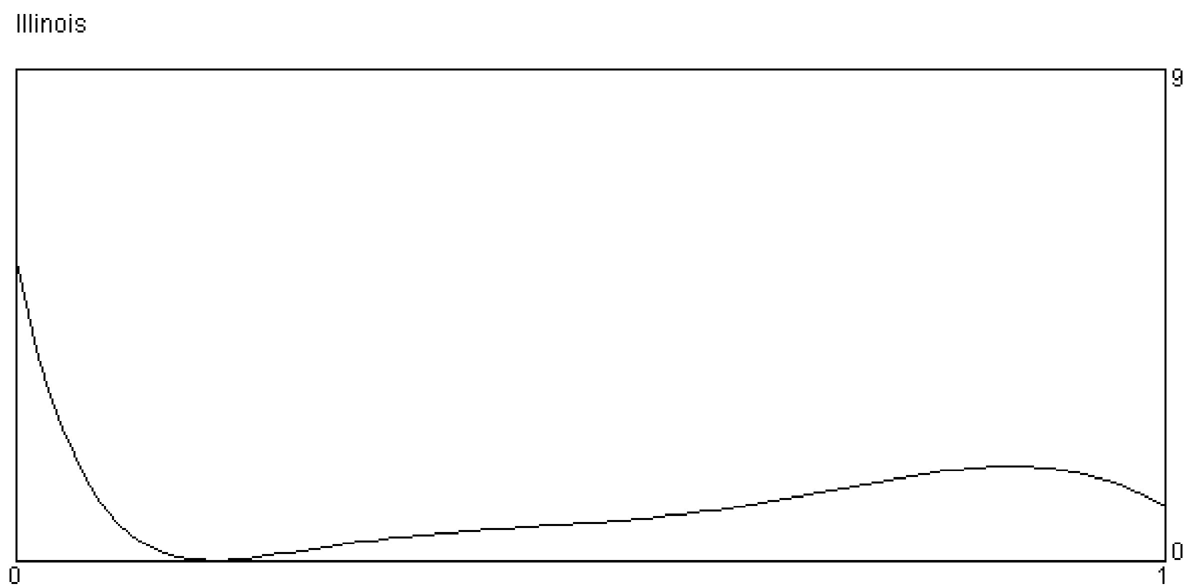

Figure 3: SNP density $h_4\left(u|\widehat{\delta}\right)$ for Illinois

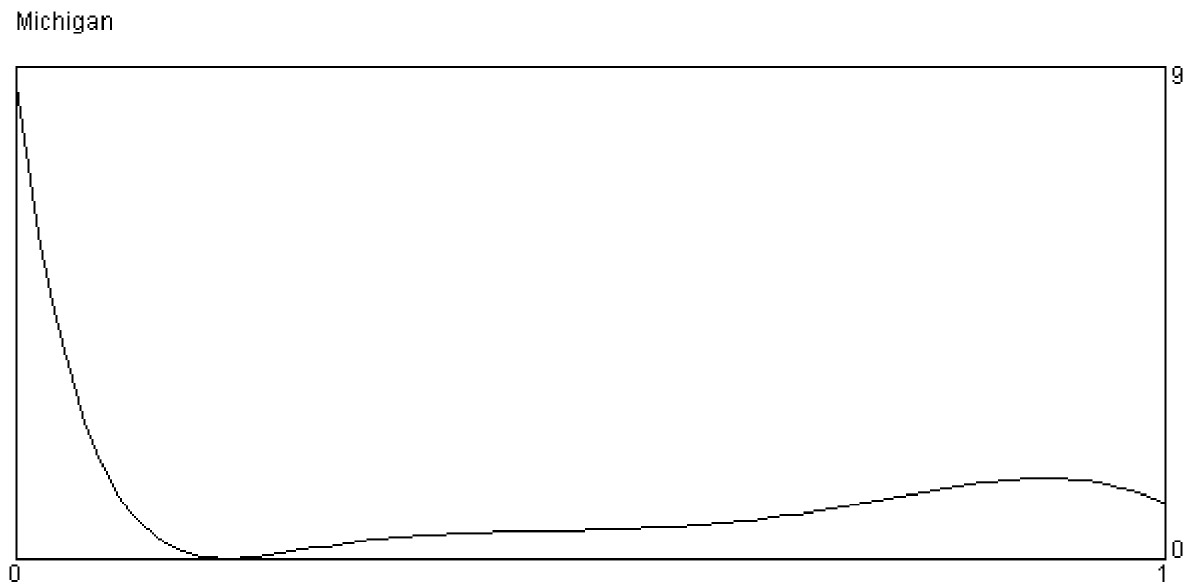

Figure 4: SNP density $h_4\left(u|\widehat{\delta}\right)$ for Michigan

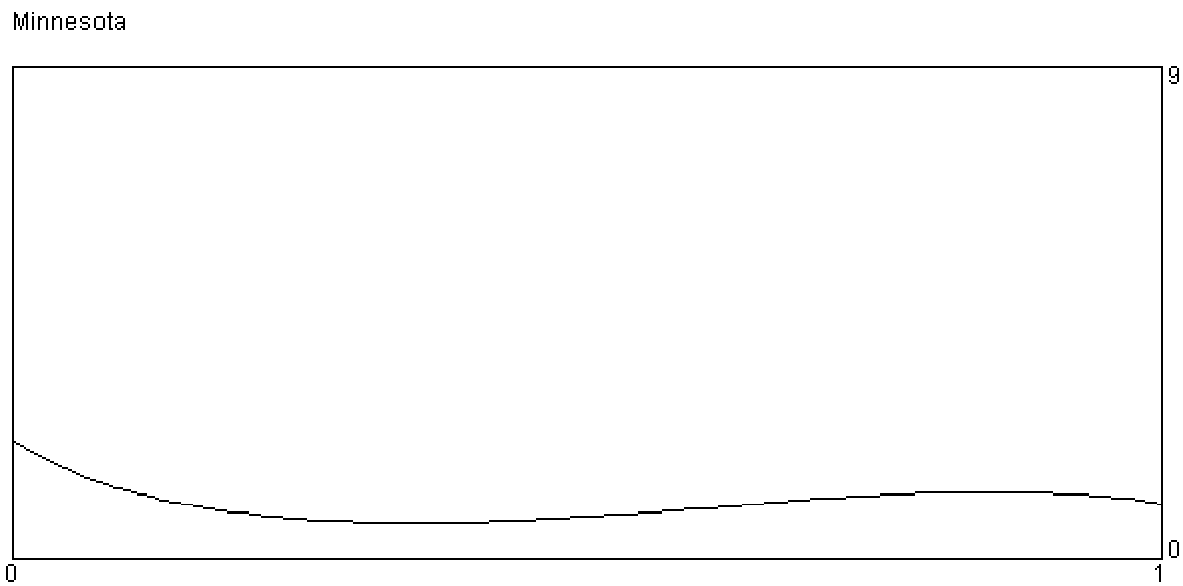

Figure 5: SNP density $h_3\left(u|\widehat{\delta}\right)$ for Minnesota

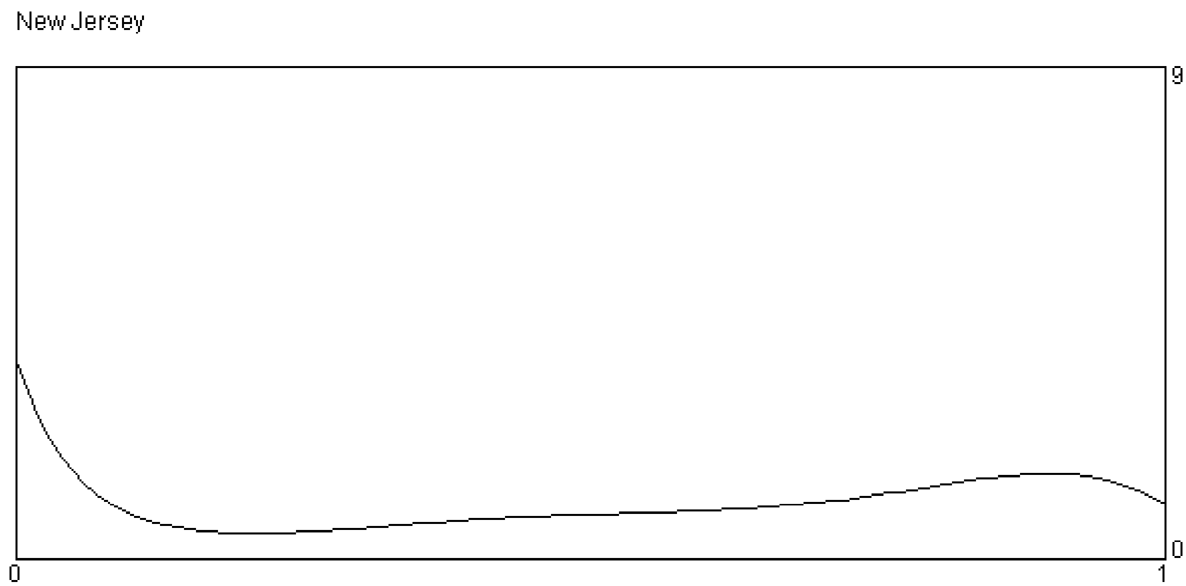

Figure 6: SNP density $h_5\left(u|\widehat{\delta}\right)$ for New Jersey

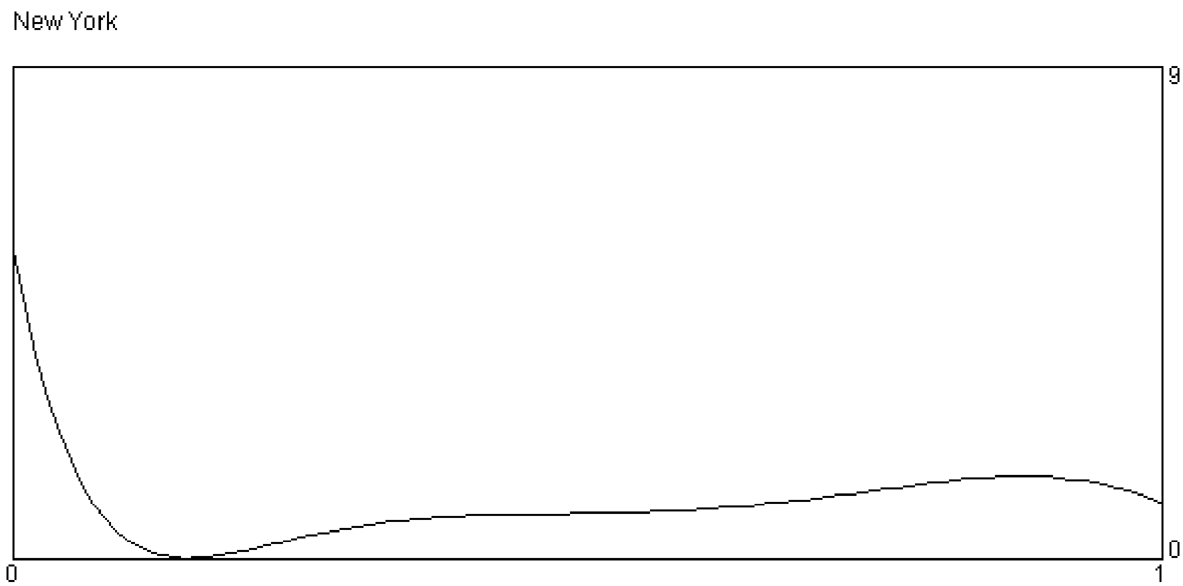

Figure 7: SNP density $h_4\left(u|\widehat{\delta}\right)$ for New York

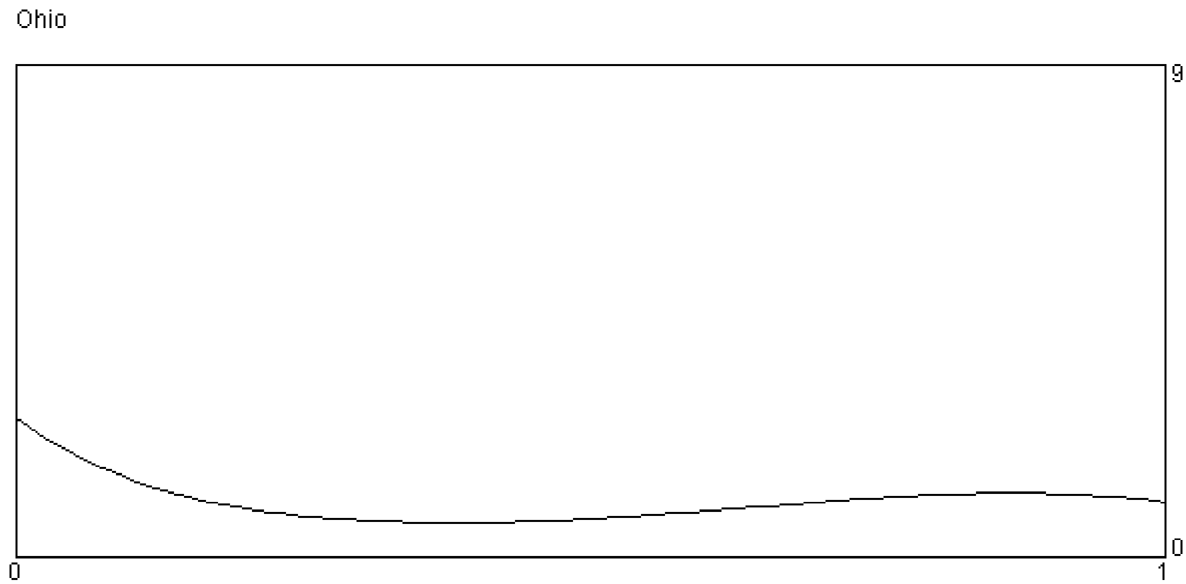

Figure 8: SNP density $h_3\left(u|\widehat{\delta}\right)$ for Ohio

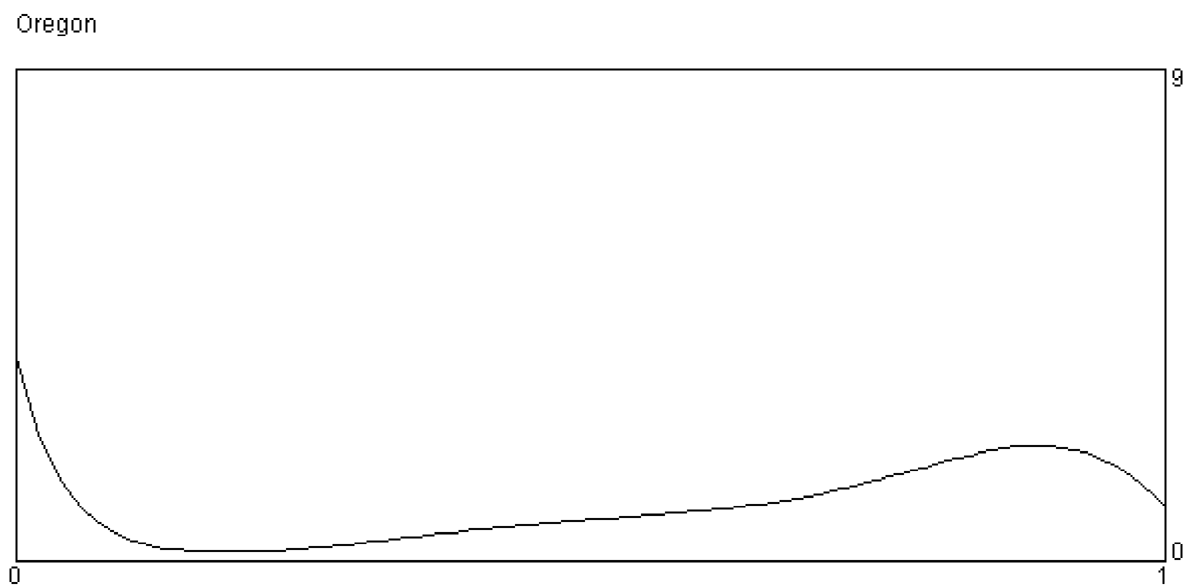

Figure 9: SNP density $h_5\left(u|\widehat{\delta}\right)$ for Oregon