Modelling Heaped Duration Data: An Application to Neonatal Mortality*

Wiji Arulampalam† Warwick University

Valentina Corradi ‡ Surrey University

Daniel Gutknecht§ Oxford University

October 22, 2014

Abstract

In 2005, the Indian Government launched a conditional cash-incentive program to encourage institutional delivery. This paper studies the effects of the program on neonatal mortality using district-level household survey data. We model mortality using survival analysis, paying special attention to the substantial heaping present in the data. The main objective of this paper is to provide a set of sufficient conditions for identification and consistent estimation of the baseline hazard accounting for heaping and unobserved heterogeneity. Our identification strategy requires neither administrative data nor multiple measurements, but a correctly reported duration and the presence of some flat segments in the baseline hazard which includes this correctly reported duration point. We establish the asymptotic properties of the maximum likelihood estimator and provide a simple procedure to test whether the policy had (uniformly) reduced mortality. While our empirical findings do not confirm the latter, they do indicate that accounting for heaping matters for the estimation of the baseline hazard.

Keywords: Discrete Time Duration Model, Heaping, Measurement Error, Parameters on the Boundary, Neonatal Mortality.

JEL Classification: C12, C21, C24, C41.

*We are grateful to Aureo de Paula, Giovanni Forchini, Federico Martellosio, and Mervin Silvapulle for useful comments and suggestions, and to Marco Alfano and Artemisa Flores for help with the data preparation.

†Department of Economics, University of Warwick, Coventry CV4 7AL, UK. Tel. +44(0)2476 523471; email: Wiji.Arulampalam@warwick.ac.uk

‡Department of Economics, University of Surrey, School of Economics, Guildford GU2 7XH, UK, Tel. +44(0)1483 693914; email: V.Corradi@surrey.ac.uk

§Corresponding Author; Department of Economics, University of Oxford, Manor Road Building, Manor Road, OX1 3UQ Oxford, UK; Nuffield College, University of Oxford, New Road, OX1 1NF Oxford, UK. Email: Daniel.Gutknecht@economics.ox.ac.uk

1 Introduction

India has one of the largest neonatal mortality and maternal mortality rates in the world.[1] Around 32 neonates per 1000 live births (around 876,200 children) die within the first month of life (Roy et al., 2013; Save the Children, 2013) and among these babies, 309,000 die on the first day. Moreover, around 200 mothers die during pregnancy and child birth per 100,000 live births. In order to tackle this huge problem, the Indian Government introduced a conditional cash-incentive (*Janani Suraksha Yojana*) program in 2005 to encourage institutional delivery. The Indian Government also deployed volunteer Accredited Social Health Activists to help mothers with antenatal and postnatal care during the crucial pre and post birth period.

This paper studies the effects of this program on neonatal mortality using district-level household survey data. We focus on the first 28 days after birth, since the effects of the program is expected to be most pronounced soon after birth when postnatal care is provided. We model mortality using survival analysis, paying special attention to a characteristic of the reported duration data which is the apparent heaping at $5, 10, 15, ...$ days, i.e. durations which are multiple of five days. One of the commonest reason for this type of heaped data is due to recall errors. Neglecting these heaping effects leads to inconsistent estimation of the hazard function (e.g. Torelli and Trivellato, 1993 and Augustin and Wolff, 2000).

In addressing these heaping effects, this paper makes a methodological contribution in the modelling of duration data when the observed data are characterized more generally by some form of abnormal concentration at certain durations. The main objective of this paper is to provide a set of sufficient conditions for identification and consistent estimation of the baseline hazard (and other model parameters) accounting for heaping and unobserved heterogeneity. We pay particular attention to the baseline hazard to gauge the effect of the policy that was specifically intended to reduce neonatal mortality.

Despite the prevalence of heaping in survey data, the econometric literature on identification and estimation of duration models with heaping is rather limited. Abrevaya and Hausman (1999) provide a set of sufficient conditions under which the monotone rank estimator is consistent for the accelerated failure time and the proportional hazard model in the presence of misreported durations. However, the object of interest in their study is to estimate the effects of the covariates and not the baseline hazard. Torelli and Trivellato (1993) on the other hand derive a likelihood function which allows some form of heaping, but require a parsimonious parametric specification for the hazard and thus their approach is not suitable to assess a policy effect on the baseline hazard. Petoussis, Gill and Zeelenberg (1997) treat heaped durations as missing values and use the Expectation-Maximization (EM) algorithm to estimate the model. Heitjan and Rubin (1990) suggest an EM-based multiple imputation method for inference in the presence of heaped data, but do not deal with duration models. Finally, Augustin and Wolff (2000) use dummy variables for heaped durations. None of these papers are interested in the identification of the baseline hazard.

The paper closest to ours is Ham, Li and Shore-Sheppard (2014). They establish identification of the baseline hazard for multiple spell durations in the presence of seam bias and unobserved heterogeneity. Seam bias, which is another form of measurement error, is characterized by the fact that the end of one spell and the beginning of the next spell do

[1] Neonatal period is the first 28 days after birth.

not coincide. The key difference between the identification strategy in Ham, Li and Shore-Sheppard (2014) and ours is that they have two measurements for a duration collected from different survey waves, where only one is affected by seam bias. By contrast, We have neither multiple measurements, nor administrative data.[2]

The identification strategy we provide is based on a set of minimal assumptions on the shape of the discretized hazard function. We require neither administrative data nor multiple measurements. The key ingredients of our identification strategy are the existence of at least one correctly reported duration and the presence of some flat segment(s) in the baseline hazard which includes this correctly reported duration point. The length of the flat segments required depends on the complexity of the heaping process. Heuristically, we use the correctly reported duration in the constant part of the baseline hazard to identify the parameters of the heaping process, i.e. the probability of rounding to a heaped value. This enables us to identify the heaping parameters and hence the rest of the baseline hazard parameters. Information about the correctly reported duration and the flat segment can stem from different sources and does not need to come from a specific data set. For instance, in the application in Section 5 we partially rely on information from a verbal autopsy report on neonatal mortality in Uttar Pradesh, which suggests that assuming a flat hazard segment towards 18 days is a relatively plausible assumption. The likelihood is constructed down-weighting the contribution of the heaped duration and over-weighting the contribution of the non heaped durations. This adjustment ensures consistent estimation of both heaping and baseline hazard parameters in the case of a parametric specification of the unobserved heterogeneity component. We rely on a parametric specification of unobserved heterogeneity to obtain a closed form for the likelihood, and to conduct inference on the baseline hazard. However, as for identification of the baseline hazard, we do emphasize that, in light of the results in Ridder and Woutersen (2003), other, more flexible choices of the heterogeneity distribution would indeed also suffice. Recently, Bierens (2008) suggests to approximate unobserved heterogeneity via Legendre polynomials, while Burda, Harding, Hausman (2014) suggest the use of an infinite mixture and, Hausman and Woutersen (2013) introduce a rank type estimator, which does not require the specification of unobserved heterogeneity. However, all these papers rule out incorrectly reported durations.

The rest of this paper is organized as follows. Section 2 describes the setup and the heaping model we consider. As a main result, it provides a set of sufficient conditions for the identification of the baseline hazard in the presence of unobserved heterogeneity and heaping. Section 3 derives the likelihood and establishes the asymptotic properties of the MLE estimator. Since we do not impose a strictly positive probability of rounding, we account for the possibility of parameters on the boundary (Andrews (1999)). We also outline (mis)specification tests for the heaping model. Section 4 investigates the effect of the cash transfer policy on the baseline hazard, taking into account a possible side effect on the heaping parameters.[3] We study whether the policy has uniformly "lowered" the baseline hazard, or if instead it had the desired effect only over part of the hazard, for example

[2]We are unable to use the durations closer to the interview as a set of 'correctly' reported observations because there is heaping among these too.

[3] In general, if mothers give birth in an institution, we would expect the birth dates to be correctly recorded.

over the first few days. Section 5 describes the data and the empirical model and reports our findings. We establish that: (i) heaping matters, in fact we find that the associated heaping probabilities before and after treatment are clearly significant; (ii) overall, the program introduction has reduced neonatal mortality, but the effect is not uniform. Section 6 concludes. Finally, Appendix I contains all technical proofs, Appendix II reports the empirical findings, while Appendix III provides details about the construction of (m out of n) bootstrap standard errors.

2 Identification of the Mixed Proportional Hazard Model with Heaped Durations

We begin by outlining our setup. We assume a Mixed Proportional Hazard (MPH) model for the unobservable true durations. Our objective is to recover the underlying structural parameters from the observable mismeasured durations.

Let τ_i^* denote the underlying "duration" of individual i for $i = 1, \ldots, N$ measured on a continuous scale. The associated hazard function for i is then given by:

$$\lambda_i(\tau^*) = \lim_{\Delta \to 0} \Pr(\tau_i^* < \tau^* + \Delta | \tau_i^* \geqslant \tau^*)/\Delta$$

We parameterise the hazard as:

$$\lambda_i(\tau^* | z_i, u_i) = \lambda_0(\tau^*) \exp(z_i' \beta + u_i),$$

where $\lambda_0(\tau^*)$ is the baseline hazard, u_i is the individual unobserved heterogeneity, and z_i a set of time invariant covariates.

We next assume that a continuous duration $\tau_i \in [\tau, \tau + 1)$ is recorded as τ. Therefore, the discrete time hazard for our model is given by:

$$\begin{aligned} h_i(\tau | z_i, u_i) =& \Pr[\tau_i < \tau + 1 | \tau_i \geqslant \tau, z_i, u_i] \\ =& 1 - \exp\left(-\int_\tau^{\tau+1} \lambda_i(s | z_i, u_i) ds\right) \\ =& 1 - \exp\left(-\exp\left(z_i' \beta + \gamma(\tau)\right) + u_i\right), \end{aligned} \tag{1}$$

where $\gamma(\tau) = \ln \int_\tau^{\tau+1} \lambda(s) ds$.

The key issue is how to identify the underlying baseline hazard when we do not observe τ_i, but only some misreported version of it. The form of misreporting we address in this paper is heaping due to rounding. Thus, in the sequel, we will provide a minimal set of conditions sufficient to identify the discretized baseline hazard in the presence of heaping.

Let $\mathcal{D}^{\mathcal{U}}$ denote the set of uncensored and $\mathcal{D}^{\mathcal{C}}$ the set of censored observations at $\overline{\tau}$, and write $\mathcal{D} = \mathcal{D}^{\mathcal{U}} \cup \mathcal{D}^{\mathcal{C}}$, i.e. for all $i = 1, .., N$, $\tau_i \in \mathcal{D}$, with $\mathcal{D} = \{0, 1, .., \overline{\tau}\}$. Our first assumption is on the censoring process.

Assumption C:

(i) Durations are censored at fixed time $\overline{\tau}$ and the censoring mechanism is independent of the durations (type I censoring; Cox and Oakes, 1984);
(ii) Censoring is independent of the heaping process.

We note that this assumption could be straightforwardly generalized to allow for varying censoring times across individuals (random censoring) as long as censoring is independent of the heaping process and C(ii) is satisfied. Also, since censoring is independent of the heaping process, deduce that $\mathcal{D}^{\mathcal{U}} = \mathcal{D}^{\mathcal{H}} \cup \mathcal{D}^{\mathcal{NH}}$, where $\mathcal{D}^{\mathcal{H}}$ and $\mathcal{D}^{\mathcal{NH}}$ are the sets of heaped and non-heaped points, respectively.[4]

In order to make our setup more formal, denote $\overline{r}$ as the maximum number of time periods that a duration can be rounded to and h^* as the first heaping point. In the following, we will assume that heaping occurs at multiples of h^*. This assumption is motivated by our application in Section 5, where reported dates of deaths are heaped at values that are multiples of 5 days (i.e. $h^* = 5$ and $\mathcal{D}^{\mathcal{H}}$ contains the durations 5, 10, 15 etc.). In addition, we also assume that the rounding is carried out to the nearest heaping point. It might be restrictive for some settings, but could easily be relaxed to non-multiple heaping points at the cost of further notation. Similarly, the symmetry in the number of time periods that people round up to or down from could be relaxed to allow for asymmetries as well.

Denote the set of
(i) heaping points as:

$$\mathcal{D}^{\mathcal{H}} = \left\{\tau : \tau = jh^*, j = 1, ..., \overline{j},\ \overline{j}h^* < \overline{\tau}\right\};$$

(ii) points that may be rounded up as:

$$\mathcal{D}^{\mathcal{H}_{-l}} = \left\{\tau : \tau = jh^* - l, j = 1, ..., \overline{j},\ \overline{j}h^* - l < \overline{\tau}\right\};$$

(iii) points that may be rounded down as:

$$\mathcal{D}^{\mathcal{H}_{+l}} = \left\{\tau : \tau = jh^* + l, j = 1, ..., \overline{j},\ \overline{j}h^* + l < \overline{\tau}\right\};$$

and
(iv) non-heaping points as:

$$\mathcal{D}^{\mathcal{NH}} = \left(\mathcal{D}^{\mathcal{C}}\right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}\right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}\right), \text{for} \quad l = 1, ..., \overline{r}.$$

Finally, all durations $\tau < \overline{\tau}$ which do not belong to $\left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}\right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}\right) \cup \mathcal{D}^{\mathcal{H}}$ lie in the complement set $\mathcal{D}^{\mathcal{T}} = \left(\left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}\right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}\right) \cup \mathcal{D}^{\mathcal{H}}\right)^c$ and are assumed to be truthfully reported.

In the following, let t_i be the potentially misreported duration and assume that if the true duration falls on one of the heaping points, it will be correctly reported. That is, for each $\tau_i \in \mathcal{D}^{\mathcal{H}}$, $t_i = \tau_i$ a.s.. However, when $\tau_i \in \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}\right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}\right)$, it is either correctly reported or rounded (up or down) to the closest heaping point belonging to $\mathcal{D}^{\mathcal{H}}$. Thus, for $l \in \{1, \ldots, \overline{r}\}$, let $\Pr(t_i = \tau_i + l) = p_l$ and $\Pr(t_i = \tau_i) = 1 - p_l$ if $\tau_i \in \mathcal{D}^{\mathcal{H}_{-l}}$.

[4] Szlydlowski (2013) allows for correlation between the censoring mechanism and unobserved heterogeneity, and, even in the absence of misreported durations, only achieves parameter set identification.

Analogously, let $\Pr(t_i = \tau_i - l) = q_l$ and $\Pr(t_i = \tau_i) = 1 - q_l$ if $\tau_i \in \mathcal{D}^{\mathcal{H}_{+l}}$. In our example, this is equivalent to assuming that a reported duration of say 5 days can include true durations of 3 and 4 (6 and 7) where they have been rounded up (down) to 5 days - ps and the qs give the probabilities of these roundings.

In order to identify the baseline hazard from possibly misreported observations, we need to put some structure on the heaping process. This is summarized in Assumption H.

Assumption H
(i) $\left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}\right) \cap \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}\right) = \emptyset$ and $\mathcal{D}^{\mathcal{H}_{+l}} \cap \mathcal{D}^{\mathcal{H}_C} = \emptyset$ for $l = \overline{r}$;
(ii) There exists $\overline{k} \leq \overline{\tau} - 2(\overline{r}+1)$, such that $\gamma(k) = \gamma(\overline{k})$ for $\overline{k} \leq k \leq \overline{k} + \overline{r} + 1$, and $\gamma(k) = \gamma(\overline{k} + \overline{r} + 2)$ for all $\overline{k} + \overline{r} + 2 \leq k < \overline{\tau}$;
(iii) $t_i = \overline{k}$ if and only if $\tau_i = \overline{k}$ a.s.;
(iv) For all $l \in \{1, \ldots, \overline{r}\}$, $p_l \in [0, 1)$ and $q_l \in [0, 1)$.

Assumption H(i) imposes that time periods cannot belong to more than one heap. This assumption, albeit restrictive, is crucial for our identification strategy and cannot be relaxed. It is, however, somewhat mitigated by the fact that we can in principle allow for a relatively complex heaping structure with differently sized heaps and rounding probabilities. H(ii)-(iii) requires that the baseline hazard is constant after time period $\overline{k}$, but possibly at different levels on either side of the heaping point $\overline{k} + \overline{r} + 1$, which could for instance apply when heaping is asymmetric.[5] Moreover, $\overline{k}$ is assumed to be observed without error, i.e. $\overline{k} \in \mathcal{D}^{\mathcal{T}}$. We emphasize that these assumptions are stronger than required as it would in principle suffice for the hazard to be constant over some region, not necessarily at the end nor even over regions that are adjacent to each other. We have made this assumption to keep the notation simple. Finally, Assumption H(iv) requires that durations belonging to either $\mathcal{D}^{\mathcal{H}_{-l}}$ or $\mathcal{D}^{\mathcal{H}_{+l}}$ have a strictly positive probability to be truthfully reported. This is an essential condition to identify $\gamma(k)$ for $1 < k < \overline{\tau}$.[6]

Heuristically, under the assumption that the hazard is constant over a set of durations which includes some truthfully reported values enables us to first uniquely identify the γs as well as the parameter modelling the heaping process, i.e. the ps and the qs, in this region. Subsequently, we can then use these ps and qs to pin down the rest of the baseline hazard parameters.

Assumption U:
(i) $v_i \equiv \exp(u_i)$ is independent of z_i;
(ii) v_i is identically and independently distributed;
(iii) The density of v is gamma with unit mean and variance σ^{-1}.
(iv) The support of at least one element z_{1i} of z_i, say $\mathcal{S}_{z1}$, whose corresponding element of β is non-zero contains at least two values. Moreover, the full support of z_i, $\mathcal{S}_z$, contains the zero vector.

[5] We note that there are different alterations of Assumption H that could identify the parameters of interest as well. For instance, dropping the assumption on $\gamma(\overline{k})$ in H(ii)-(iii), one could still obtain the result of Proposition 1 below if the γ parameters were constant *and* the same to the right and the left of the heaping point (rather than to differ in their levels as in H(ii)).

[6] One could also assume that the ps and qs are a function of some characteristics if we impose further structure.

Assumption U(i)-(ii) allows to integrate out unobserved heterogeneity and so to identify the unconditional hazard function. The parametric choice of the unobserved heterogeneity distribution in Assumption U(iii) on the other hand allows to obtain a closed form expression for the unconditional hazard function, which will be used in the identification proof of Proposition 1 below. In fact, identification of the baseline hazard together with the ps and the qs would only require some mild regularity conditions as in Ridder and Woutersen (2003). While the gamma density choice might appear overly restrictive at first sight, we note that U(iii) can often be rationalised theoretically (Abbring and Van Den Berg, 2007) and findings by Han and Hausman (1990) as well as Meyer (1990) suggest that estimation results for discrete-time proportional hazard models where the baseline is left unspecified (as in our model) display little sensitivity to alternative distributional assumptions on v_i. Finally, albeit beyond the scope of this paper, it might be possible to adopt other, more flexible approaches such as the one recently proposed by Burda, Harding, and Hausman (2014). Finally, Assumption U (iv) is standard in the literature on identification of MPH models (cf. Elbers and Ridder, 1982; Ridder and Woutersen, 1984) and requires a minimum amount of variation in the covariates z_i to identify β.

Before stating our main identification result, we need to define some more notation, which will be used in the proof of Proposition 1 below. Let $\underline{\theta} = \left(\beta, \gamma(0), \gamma(1), ..., \gamma(\overline{k} - 1), \gamma(\overline{k}), \gamma(\overline{k} + \overline{\tau} + 2)\right)$ and define the probability of survival at least until time d in the absence of misreporting as:

$$\begin{aligned}
S_i\left(d|z_i, u_i, \underline{\theta}\right) &= \Pr\left(t_i \geq d|z_i, u_i, \underline{\theta}\right) \\
&= \prod_{s=0}^{d-1} \exp\left(-\exp\left(z_i'\beta + \gamma(s) + u_i\right)\right) \\
&= \prod_{s=0}^{d-1} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s)\right)\right).
\end{aligned}$$

Now, for durations that are censored at time period $\overline{\tau}$ we have:

$$\begin{aligned}
S_i\left(\overline{\tau}|z_i, u_i, \underline{\theta}\right) &= \Pr\left(t_i \geq \overline{\tau}|z_i, u_i, \underline{\theta}\right) \\
&= \prod_{s=0}^{\overline{\tau}-1} \exp\left(-\exp\left(z_i'\beta + \gamma(s) + u_i\right)\right) \\
&= \prod_{s=0}^{\overline{\tau}-1} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s)\right)\right).
\end{aligned}$$

Moreover, the probability for an exit event in $t_i < \overline{\tau}$ is:

$$\begin{aligned}
f_i\left(\tau|z_i, u_i, \underline{\theta}\right) &= \Pr\left(t_i = \tau|z_i, u_i, \underline{\theta}\right) \\
&= S_i\left(\tau|z_i, u_i, \underline{\theta}\right) - S_i\left(\tau + 1|z_i, u_i, \underline{\theta}\right) \\
&= \prod_{s=0}^{\tau-1} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s)\right)\right) \\
&\quad - \prod_{s=0}^{\tau} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s)\right)\right).
\end{aligned} \tag{2}$$

Here, $f_i(\tau|z_i, u_i, \underline{\theta})$ denotes the probability of a duration equal to τ when there is no misreporting. However, because of the rounding, heaped values are over-reported while non-heaped values are under-reported, and this needs to be taken into account when constructing the likelihood (see next section). Hereafter, let

$$\phi_i(t|z_i, v_i, \underline{\theta}) = \Pr(t_i = t|z_i, v_i, \underline{\theta})$$

with t denoting the discrete *reported* duration. It is immediate to see that

(i) for $t_i \in \mathcal{D}^{\mathcal{T}}$, $\quad \phi_i(t|z_i, v_i, \underline{\theta}) = f_i(t|z_i, v_i, \underline{\theta})$;

(ii) for $t_i \in \mathcal{D}^{\mathcal{C}}$, $\quad \Pr(t_i \geqslant t|z_i, v_i, \underline{\theta}) = S_i(t|z_i, v_i, \underline{\theta})$, with $t = \overline{\tau}$;

(iii) for $t_i \in \mathcal{D}^{\mathcal{H}_{-l}}$, $\quad \phi_i(t|z_i, v_i, \underline{\theta}) = (1 - p_l) f_i(t|z_i, v_i, \underline{\theta})$;

(iv) for $t_i \in \mathcal{D}^{\mathcal{H}_{+l}}$, $\quad \phi_i(t|z_i, v_i, \underline{\theta}) = (1 - q_l) f_i(t|z_i, v_i, \underline{\theta})$;

(v) and for $t_i \in \mathcal{D}^{\mathcal{H}}$,

$$\phi_i(t|z_i, v_i, \underline{\theta}) = \sum_{l=1}^{\overline{r}} p_l f_i(t - l|z_i, v_i, \underline{\theta}) + \sum_{l=1}^{\overline{r}} q_l f_i(t + l|z_i, v_i, \underline{\theta}) + f_i(t|z_i, v_i, \theta).$$

In summary, there are five different probabilities of exit events depending on whether the reported duration t_i is in $\mathcal{D}^{\mathcal{T}}$, $\mathcal{D}^{\mathcal{C}}$, $\mathcal{D}^{\mathcal{H}_{-l}}$, $\mathcal{D}^{\mathcal{H}_{+l}}$, or $\mathcal{D}^{\mathcal{H}}$ respectively.

Moreover, using assumption U, the unconditional probabilities in case (i) above are given by:

$$\begin{aligned}
\int \phi_i(t|z_i, v, \underline{\theta})\, g(v;\sigma) dv &= \int f_i(t|z_i, v, \underline{\theta})\, g(v;\sigma) dv \\
&= \int \Pr(\tau_i = t|z_i, v, \underline{\theta})\, g(v;\sigma) dv \\
&= \int \left[S_i(t|z_i, v, \underline{\theta}) - S_i(t+1|z_i, v, \underline{\theta})\right] g(v;\sigma) dv \\
&= \int S_i(t|z_i, v, \underline{\theta})\, g(v;\sigma) dv - \int S_i(t+1|z_i, v, \underline{\theta})\, g(v;\sigma) dv \\
&= \left(1 + \sigma\left(\sum_{s=0}^{t-1} \exp\left(z_i'\beta + \gamma(s)\right)\right)\right)^{-\sigma^{-1}} \\
&\quad - \left(1 + \sigma\left(\sum_{s=0}^{t} \exp\left(z_i'\beta + \gamma(s)\right)\right)\right)^{-\sigma^{-1}}
\end{aligned}$$

and in case (ii) by:

$$\begin{aligned}
\int \Pr(t_i \geqslant t|z_i, v, \underline{\theta})\, g(v;\sigma) dv &= \int S_i(t|z_i, v, \underline{\theta})\, g(v;\sigma) dv \\
&= \left(1 + \sigma\left(\sum_{s=0}^{t-1} \exp\left(z_i'\beta + \gamma(s)\right)\right)\right)^{-\sigma^{-1}},
\end{aligned}$$

where the last equalities use the fact that there is a closed form expression for the case of a Gamma density in a Proportional Hazard model (e.g., see Meyer (1990, p. 770)). Moreover, since the integral is a linear operator the probabilities for the cases (iii) to (v) can be derived accordingly.

Proposition 1: *Given Assumptions* C, H, *and* U, *we can uniquely identify the baseline hazard parameters* $\gamma(0)$ *to* $\gamma(\overline{\tau}-1)$ *together with the heaping probabilities* p_l *and* q_l *for* $l = \{1, \ldots, \overline{r}\}$ *from the reported durations.*

The proof is based on establishing a one to one relationship between survival probabilities and the baseline hazard parameters. Given Assumptions U, H(ii)-(iii), we follow the same approach as in Heckman and Singer (1984) to uniquely identify $\sum_{s=0}^{\overline{k}} \exp(\gamma(s))$ and so $\gamma(\overline{k})$. Given this, and exploiting the flatness of the hazard, as stated in H(ii), we identify the heaping probabilities ps and qs. Finally, using H(iv), we sequentially identify all $\gamma(s)$, for $s < \overline{k}$.[7]

3 Estimation of the Mixed Proportional Hazard Model with Heaped Durations.

Our next goal is to obtain consistent estimators for $\theta = \{\underline{\theta}, \sigma\}$ from the possibly misreported durations. To do this, we first set up the likelihood function drawing from the derivations of the previous section for truthfully and misreported durations. That is, given Assumption U and the definition of $\phi_i(\cdot)$ from cases (i) to (v), let:

$$L_N(\theta) = \prod_{i=1}^{N} \int \phi_i(t|z_i, v) g(v; \sigma) \mathrm{d}v$$

and so

$$\begin{aligned} l_N(\theta) &= \ln L_N(\theta) \\ &= \sum_{i=1}^{N} \ln \int \phi_i(t|z_i, v) g(v; \sigma) \mathrm{d}v. \end{aligned}$$

Thus

$$\widehat{\theta}_N = \arg\max_{\theta \in \Theta} l_N(\theta)$$

$$\theta^{\dagger} = \arg\max_{\theta \in \Theta} \mathrm{E}\left(l_N(\theta)\right).$$

Assumption D:
(i) Assume that $\mathbb{E}[\tau_i^4] < \infty$.
(ii) The durations τ_i, $i = 1, ..., N$ are identically and independently distributed.

[7] Note that, similar to Assumption H, it appears from the proof in the Appendix that Assumption U is sufficient, but by no means necessary.

(iii) For all $d = 1, ..., \overline{\tau}$, $\frac{1}{N} \sum_i 1\{\tau_i = d\} \xrightarrow{p} \Pr[\tau_i = d] > 0$.

Together with Assumption H(iv), Assumption D(iii) ensures that we observe exits in each time period until $\overline{\tau}$. Needless to say, if we do not have enough observations for a given duration, we cannot consistently estimate the associated baseline hazard parameters. We now establish the asymptotic properties of $\widehat{\theta}_N$.

Theorem 2: *Let Assumptions* H,U,C *and* D *hold. Then:*
(i)

$$\widehat{\theta}_N - \theta^\dagger = o_p(1)$$

(ii)

$$\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right) \xrightarrow{d} \inf_{\psi \in \Psi} \left((\psi - G)' \mathcal{I}^\dagger (\psi - G) \right),$$

with $\mathcal{I}^\dagger = E\left(\left(-\nabla^2_{\theta\theta} l_N(\theta)/N\right)|_{\theta=\theta^\dagger}\right)$, *and* $G \sim N\left(0, \mathcal{I}^{\dagger -1}\right)$, Ψ *being a cone in* $R^{p_\beta + \overline{k} + 2 + 2\overline{r}}$.[8]

(iii) Let $\pi^\dagger = \left(p_1^\dagger, ..., p_{\overline{r}}^\dagger, q_1^\dagger, ..., q_{\overline{r}}^\dagger\right)'$, *if* $\pi^\dagger \in (0,1)^{2\overline{r}}$, *then*

$$\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right) \xrightarrow{d} N\left(0, \mathcal{I}^{\dagger -1}\right).$$

In the current context, we are particularly interested in carrying out inference on the baseline hazard parameters, and for that we use critical values from the limiting distribution of $\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right)$. However, as the information matrix $\mathcal{I}^{\dagger -1}$ is not block diagonal, the limiting distribution of the baseline hazard parameters depends on whether some heaping probabilities are equal to zero or not. If one or more of the "true" rounding probabilities are equal to zero, then the limiting distribution of $\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right)$ is no longer normal. Needless to say, this complicates inference on the baseline hazard.

Thus, we want to test the null hypothesis that at least one rounding parameter is equal to zero versus the alternative that none is zero. If we reject the null, then we know that we do not have any boundary problem and can then rely on the asymptotic normality result in Theorem 2(iii). Let $H_{p,0}^{(j)} : p_j = 0$, $H_{p,A}^{(j)} : p_j > 0$ and let $H_{q,0}^{(j)}, H_{q,A}^{(j)}$ be defined analogously. Our objective is to test the following hypotheses,

$$H_0^{\overline{r}} = \left(\cup_{j=1}^{\overline{r}} H_{p,0}^{(j)}\right) \cup \left(\cup_{j=1}^{\overline{r}} H_{q,0}^{(j)}\right)$$

vs

$$H_A^{\overline{r}} = \left(\cap_{j=1}^{\overline{r}} H_{p,A}^{(j)}\right) \cap \left(\cap_{j=1}^{\overline{r}} H_{q,A}^{(j)}\right),$$

so that under $H_A^{\overline{r}}$ all ps and qs are strictly positive.

To decide between $H_0^{\overline{r}}$ and $H_A^{\overline{r}}$ we follow the Intersection-Union principle, IUP, see e.g. Chapter 5 in Silvapulle and Sen (2005). According to the IUP, we reject $H_0^{\overline{r}}$ at level α, if

[8] Ψ is a cone in $\mathbb{R}^s$, if for $a > 0$, $\psi \in \Psi$ implies $a\psi \in \Psi$. Moreover, p_β denotes the dimension of z_i.

all single null hypotheses $H_{p,0}^{(j)}$ and $H_{q,0}^{(j)}$ are rejected at level α. Otherwise, we do not reject $H_0^{\overline{r}}$.

Let

$$t_{p_j,N} = \left(\widehat{\mathcal{I}}_{p_jp_j,N}^{1/2}\right)\widehat{p}_{j,N}, \qquad t_{q_j,N} = \left(\widehat{\mathcal{I}}_{q_jq_j,N}^{1/2}\right)\widehat{q}_{j,N},$$

where $\widehat{\mathcal{I}}_N^{1/2}\widehat{\mathcal{I}}_N^{1/2} = \widehat{\mathcal{I}}_N$, $\widehat{\mathcal{I}}_N = \frac{1}{N}\nabla_\theta l_N(\theta^\dagger)\nabla'_\theta l_N(\theta^\dagger)$ and $\widehat{\mathcal{I}}_{p_jp_j,N}, \widehat{\mathcal{I}}_{q_jq_j,N}$ are the corresponding entries.[9] Also, let

$$PV_{p,j,N} = \Pr\left(Z > t_{p_j}\right), \quad PV_{q,j,N} = \Pr\left(Z > t_{q_j}\right),$$

where Z denotes a standard normal random variable.

We now introduce a rule for deciding between $H_0^{\overline{r}}$ and $H_A^{\overline{r}}$.

Rule IUP-PQ: Reject $H_0^{\overline{r}}$, if $\max_{j=1,\dots,\overline{r}}\left\{PV_{p,j}, PV_{q,j}\right\} < \alpha$ and do not reject otherwise.

Proposition 3: *Let Assumptions* H,U,C *and* D *hold.* Then, Rule IUP-PQ ensures that

$$\lim_{N\to\infty}\Pr\left(\text{Reject } H_0^{\overline{r}}|H_0^{\overline{r}} \text{ true}\right) \leq \alpha$$

$$\lim_{N\to\infty}\Pr\left(\text{Reject } H_0^{\overline{r}}|H_0^{\overline{r}} \text{ false}\right) = 1.$$

If we reject $H_0^{\overline{r}}$, we can proceed performing inference based on asymptotic normality. If we fail to reject $H_0^{\overline{r}}$, then we drop the p_j or q_j with the largest associated p-value and we apply Rule IUP once again.

4 Modeling the Effect of the Cash Transfer Policy

Our main empirical question is to assess whether the introduction of the cash transfer policy has reduced neonatal mortality. More precisely, we want to device a procedure for testing the hypothesis that the baseline hazard has been lowered by the policy. In order to capture the effect of the cash transfer policy we introduce a dummy D_i, where $D_i = 1$ if the duration measurements started after the introduction of the policy, and zero otherwise. The discrete hazard function for the true duration is

$$\begin{aligned}\widetilde{h}_i(d|z_i,u_i) &= \Pr[\tau_i < d+1|\tau_i \geq d, z_i, u_i] \\ &= \left(1-\exp\left(-\exp\left(z_i'\beta + \gamma(k) + \gamma^{(2)}(k)D_i + u_i\right)\right)\right),\end{aligned}$$

where the coefficient of D_i, $\gamma^{(2)}(k)$, is defined analogously to $\gamma(k)$. It is immediate to see that $\gamma^{(2)}(k) < 0$ implies a lower hazard after the policy introduction.

We want to isolate any possible confounding effect. For example, it might be the case that the heaping probabilities are also affected by the program: if more women deliver babies in hospitals and are also followed up after the birth after the implementation of

[9] From (2) and the definition of $\phi_i(\cdot)$ in (i) to (v), it is immediate to see that we can take right and left derivatives and evaluate at the boundary level. See discussion in the proof of Theorem 2(ii) in the Appendix.

the program, births and deaths might, on average, be recorded more accurately than before. That is, if more women deliver in hospitals, then it is likely that the probability of rounding will decrease as families typically receive birth certificates when being discharged from the hospital, which allows them to recall more accurately. Therefore, to isolate the genuine effect on neonatal mortality, we allow the rounding probabilities to vary after the policy introduction. For all durations not truthfully reported we allow for possibly different rounding errors depending on whether the reported duration occurred before or after the policy introduction.[10]

More formally, let $\underline{\vartheta} = \{\underline{\theta}, \gamma^{(2)}(0), \ldots, \gamma^{(2)}(\overline{k}-1), \gamma^{(2)}(\overline{k}), \gamma^{(2)}(\overline{k}+\overline{r}+2)\}$ and $\vartheta = \{\underline{\theta}, \sigma\}$. Define the likelihood contribution of a correctly reported duration as:

$$\begin{aligned}\widetilde{f}_i\left(\tau|z_i, u_i, \underline{\vartheta}\right) &= \Pr\left(t_i = \tau|z_i, u_i, \underline{\vartheta}\right) \\ &= \prod_{s=0}^{\tau-1} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s) + \gamma^{(2)}(s)D_i\right)\right) \\ &\quad - \prod_{s=0}^{\tau} \exp\left(-v_i \exp\left(z_i'\beta + \gamma(s) + \gamma^{(2)}(s)D_i\right)\right).\end{aligned}$$

Then, the contribution of a non-truthfully reported duration can be defined in analogy to $\phi_i\left(t|z_i, u_i, \underline{\theta}\right)$, say $\widetilde{\phi}_i\left(t|z_i, u_i, \underline{\vartheta}\right)$. Thus,
(i) for any $t_i = t \in \mathcal{D}^{\mathcal{H}_{-l}}$,

$$\widetilde{\phi}_i(k|z_i, v_i, \underline{\vartheta}) = (1 - p_l^{(1)}(1 - D_i) - p_l^{(2)} D_i)\widetilde{f}_i(k|z_i, v_i, \underline{\vartheta}),$$

(ii) for $t_i = t \in \mathcal{D}^{\mathcal{H}_{+l}}$

$$\widetilde{\phi}_i(k|z_i, v_i, \underline{\vartheta}) = (1 - q_l^{(1)}(1 - D_i) - q_l^{(2)} D_i)\widetilde{f}_i(k|z_i, u_i, \underline{\vartheta}),$$

(iii) and for $t_i = t \in \mathcal{D}^{\mathcal{H}}$,

$$\begin{aligned}\widetilde{\phi}_i(k|z_i, u_i, \underline{\vartheta}) &= \sum_{l=1}^{\overline{r}} (p_l^{(1)}(1 - D_i) + p_l^{(2)} D_i)\widetilde{f}(k - l|z_i, u_i, \underline{\vartheta}) \\ &\quad + \sum_{l'=1}^{\overline{r}} (q_l^{(1)}(1 - D_i) + q_l^{(2)} D_i)\widetilde{f}_i(k + l|z_i, u_i, \underline{\vartheta}) \\ &\quad + \widetilde{f}(k|z_i, u_i, \underline{\vartheta}).\end{aligned}$$

Notice that the specification above allows for different heaping probabilities before and after the introduction of the policy.
Finally, let $\widetilde{l}_N(\vartheta)$ be defined as $l_N(\theta)$, but with $\widetilde{\phi}_i\left(t|z_i, u_i, \underline{\vartheta}\right)$ instead of $\phi_i\left(t|z_i, u_i, \underline{\theta}\right)$. Also let

$$\widetilde{\vartheta}_N = \arg\max_{\vartheta \in \widetilde{\Theta}} \widetilde{l}_N(\vartheta)$$

[10]Note, however, that our setup does not allow the β's to change after the program introduction.

$$\vartheta^{\ddagger} = \arg\max_{\vartheta\in\widetilde{\Theta}} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\left(\widetilde{l}_N(\vartheta)\right).$$

Note that durations are not necessarily identically distributed, as they may differ depending on whether they occur before or after treatment.

We formulate the null as

$$H_0^{\gamma} : \max\left\{\gamma^{(2)}(0), \gamma^{(2)}(1),..., \gamma^{(2)}(\overline{k}), \gamma^{(2)}(\overline{k}+\overline{r}+2)\right\} \geq 0$$

versus

$$H_A^{\gamma} : \max\left\{\gamma^{(2)}(0), \gamma^{(2)}(1),..., \gamma^{(2)}(\overline{k}), \gamma^{(2)}(\overline{k}+\overline{r}+2)\right\} < 0.$$

The null hypothesis is that over at least one "day" the hazard function either has increased or has not changed. On the other hand, under the alternative, the policy has reduced neonatal mortality over the all period considered, i.e. over every day the baseline hazard has decreased. Note that H_A^{γ} implies

$$\widetilde{H}_A^{\gamma} : \max_{J\leq\overline{k}+\overline{r}+2}\left\{\sum_{j=0}^{J}\gamma^{(2)}(j)\right\} < 0$$

while $\widetilde{H}_A^{\gamma}$ does not necessarily imply H_A^{γ}. Thus, rejection of H_0^{γ} is a sufficient, but not a necessary condition, for a uniform shift upward of the survivor function. In other words, if we reject the null we have strong evidence that the policy has generated the desired effect. Now, with a slight abuse of notation, it is immediate to see that we can re-state H_0^{γ} and H_A^{γ} as,

$$H_0^{\gamma} = \cup_{j=1}^{\overline{k}+2} H_{\gamma,0}^{(j)}$$

vs

$$H_A^{\gamma} = \cap_{j=1}^{\overline{k}+2} H_{\gamma,A}^{(j)},$$

where $H_{\gamma,0}^{(j)} : \gamma^{(2)}(j) \geq 0$ and $H_{\gamma,A}^{(j)} : \gamma^{(2)}(j) < 0$. Thus, the null implies that for at least one j, $\gamma^{(2)}(j) \geq 0$ while the alternative is that $\gamma^{(2)}(j) < 0$ for all j. Thus we can apply again the Intersection Union Principle, IUP. Let:

$$t_{\gamma_j^{(2)},N} = \left(\widehat{\mathcal{I}}^{1/2}_{\gamma_j^{(2)}\gamma_j^{(2)},N}\right)\widetilde{\gamma}_{j,N}^{(2)}, \quad PV_{\gamma^{(2)},j,N} = \Pr\left(Z > t_{\gamma_j^{(2)},N}\right),$$

with Z being a standard normal random variable.

Rule IUP-GAMMA2: Reject H_0^{γ}, if $\max_{j=1,...,\overline{k}+2}\left\{PV_{\gamma^{(2)},j,N}\right\} < \alpha$ and do not reject otherwise.

In the sequel we shall need

Assumption D':
(i) Assume that $\mathbb{E}[\tau_i^{4(1+\delta)}] < \infty$ for $\delta > 0$.
(ii) The duration τ_i, $i = 1, ..., N$ are independently but not identically distributed.
(iii) As in Assumption D.

As mentioned above, durations are no longer identically distributed because of the possible structural break due to the policy introduction.

Proposition 4: *Let Assumptions* H,U,C *and* D' *hold. Then, Rule IUP-GAMMA2 ensures that*

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\gamma} | H_0^{\gamma} \text{ true}\right) \leq \alpha$$

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\gamma} | H_0^{\gamma} \text{ false}\right) = 1.$$

If we reject H_0^{γ}, we can stop. In fact, its rejection provides strong evidence for the efficacy of the policy. If we instead fail to reject H_0^{γ}, the natural step is to proceed to test the null hypothesis that the introduction of the cash transfer has not decreased the probability of a baby dying in any of the first $\overline{\tau}$ days, against the alternative that over at least one day the probability of death has decreased. Hence, formally if we fail to reject H_0^{γ}, we proceed to test

$$H2_0^{\gamma} : \min\left\{\gamma^{(2)}(0), \gamma^{(2)}(1), ..., \gamma^{(2)}(\overline{k} + \overline{r} + 2)\right\} \geq 0$$

versus

$$H2_A^{\gamma} : \min\left\{\gamma^{(2)}(0), \gamma^{(2)}(1), ..., \gamma^{(2)}(\overline{k} + \overline{r} + 2)\right\} < 0.$$

Thus the null is that the introduction of the cash transfer has not decreased the probability of a child dying in any of the first $\overline{\tau}$ days. The alternative is that over at least one day the probability of death has decreased. Note that $H2_0^{\gamma}$ implies

$$\widetilde{H}2_0^{\gamma} : \min_{J \leq \overline{k} + \overline{r} + 2} \left\{\sum_{j=0}^{J} \gamma^{(2)}(j)\right\} \geq 0$$

while $\widetilde{H}2_0^{\gamma}$ does not necessarily imply $H2_0^{\gamma}$. Thus, failure to reject $\widetilde{H}2_0^{\gamma}$ is a sufficient, but not necessary condition, for a uniform shift downward or for no change in the survivor function. The null is a composite hypothesis, as it is equivalent to (again, with a slight abuse of notation)

$$H2_0^{\gamma} = \cap_{j=1}^{\overline{k}+2} H2_{0,j}^{(\gamma)}$$

where $H2_{0,j}^{(\gamma)} : \gamma_j^{(2)} \geq 0$. At issue here is the control of the overall size when testing composite hypotheses.[11] One common approach to this problem is based on controlling the overall Family-Wise Error-Rate (FWER), which ensures that no single hypothesis is rejected at a level larger than a fixed value, say α. This is typically accomplished by sorting individual $p-$values, and using a rejection rule which depends on the overall number of hypotheses. For further discussion, see Holm (1979), who develops modified Bonferroni bounds, White (2000), who develops the so-called "reality check", and Romano and Wolff (2005), who provide a refinement of the reality check, in terms of controlling the average number of false nulls rejected. However, when the number of hypotheses in the composite is large, all these procedures tend to be rather conservative. This is because, strictly positive elements of

[11] For a survey of recent developments on testing composite hypotheses, see Corradi and Distaso (2011).

$\gamma^{(2)}$ do not contribute to the statistic, but do contribute to the p-vale values, making them larger.

A less conservative approach can be based on Andrews and Soares (2010) Generalized Moment Selection (GMS). Consider the following statistic:[12]

$$S_N^- = \sum_{j=0}^{\overline{k}+1} \left(\sqrt{N} \widetilde{\gamma}_N^{(2)}(j) / \widetilde{\sigma}_{N,j,j} \right)_-^2, \text{ where } x_- = \begin{cases} x, & \text{if } x \leq 0 \\ 0, & \text{if } x > 0 \end{cases}, \tag{3}$$

where $\widetilde{\sigma}^2_{N,j,j}$ is the jj diagonal element of $R\widetilde{\Sigma}_N R'$, with $\widetilde{\Sigma}_N$ being an estimator of vcov$\left(\sqrt{N}\left(\widetilde{\vartheta}_N - \vartheta^{\ddagger}\right)\right)$, and $R\widetilde{\theta}_N = \widetilde{\gamma}^{(2)} = \left(\widetilde{\gamma}^{(2)}(0), ..., \widetilde{\gamma}^{(2)}(\overline{k} + \overline{r} + 2)\right)$. Notice that, when $\widehat{\gamma}^{(2)}(k) \geq 0$ for all k, S_N^- is equal to zero almost surely. In fact, only negative elements of $\widehat{\gamma}_2$ contribute to the statistic.

Theorem 5: *Let Assumptions H,U,C and D' hold. Then, under* $H2_0^{\gamma}$,

$$S_N^- \xrightarrow{d} \sum_{j=0}^{\overline{k}+1} \left(\sum_{i=0}^{\overline{k}+1} \omega_{ji} Z_i + h_j \right)_-^2,$$

where $Z \sim N(0, I_{\overline{k}+2})$, ω^2_{ji} *is the* j, i *element of* $\Omega^{1/2}$, *and* $\Omega = D^{-1/2}\left(R\Sigma R'\right) D^{-1/2}$, *with* $D = diag\left(R\Sigma R'\right)$, *and* $h_j = \lim_{N\to\infty} \left(\sqrt{N}\gamma^{(2)}(j)\right)$. *Under* $H2_A^{\gamma}$, *for* $\varepsilon > 0$,

$$\lim_{N\to\infty} \Pr\left(\frac{1}{\sqrt{N}} S_N^- > \varepsilon \right) = 1.$$

In order to construct valid critical values for S_N^-, we can easily simulate variates from $\widetilde{\Omega}_N^{1/2} N\left(0, I_{\overline{k}+2}\right)$, but we need a way of approximating the slackness vector h. The main problem is that the vector h cannot be consistently estimated. Intuitively, except for the least favorable case under the null, i.e. the case of $\gamma^{(2)}(0) = ... = \gamma^{(2)}\left(\overline{k}\right) = \gamma^{(2)}\left(\overline{k} + \overline{r} + 2\right) = 0$, $\left(\gamma^{(2)}(k)/\omega_{k,k}\right) > 0$ and so $\lim_{N\to\infty} \sqrt{N}\left(\gamma^{(2)}(k)/\omega_{k,k}\right)$ tends to infinity, and cannot be consistently estimated. The idea behind the GMS approach is to define data-driven rules to approximate h and control for the degree of slackness. In the sequel, we choose h according to the following rule, based on the law of the iterated logarithm: if $\widetilde{\gamma}_N^{(2)}(j) \leq \widetilde{\sigma}_{N,j,j}\sqrt{2\ln(\ln(N))/N}$, then $h_j = 0$ otherwise if $\widetilde{\gamma}_N^{(2)}(j) > \widetilde{\sigma}_{N,j,j}\sqrt{2\ln(\ln(N))/N}$, then $h_j = \infty$, so that $\sum_{i=0}^{\overline{k}+1} \omega_{ji} Z_i + h_j > 0$ almost surely and thus it does not contribute to the computation of the simulated critical values. Hereafter, let $c^*_{B,N,(1-\alpha)}$ be the $(1-\alpha)-$th percentile of the empirical distribution of

$$S_N^{-*(b)} \xrightarrow{d} \sum_{j=0}^{\overline{k}+1} \left(\sum_{i=0}^{\overline{k}+1} \widetilde{\omega}_{N,ji} \eta_i^{(b)} 1\left\{ \widetilde{\gamma}_N^{(2)}(j) \leq \widetilde{\sigma}_{N,j,j}\sqrt{2\ln(\ln(N))/N} \right\} \right)_-^2,$$

where for $b = 1, ..., B$, $\eta^{(b)} = \left(\eta_0^{(b)}, ..., \eta_{\overline{k}+1}^{(b)}\right)' \equiv N\left(0, I_{\overline{k}+1}\right)$.

[12] S_N^- is the same as the criterion function in Chernozukov, Hong and Tamer (2007).

Theorem 6: *Let Assumptions H,U,C and D' hold. Then, under $H2_0^\gamma$,*

$$\lim_{N,B\to\infty} \Pr\left(S_N^- \leq c^*_{B,N,(1-\alpha)}\right) \geq 1-\alpha. \tag{4}$$

and under $H2_A^\gamma$,

$$\lim_{N,B\to\infty} \Pr\left(S_N^- \leq c^*_{B,N,(1-\alpha)}\right) = 0. \tag{5}$$

Note that none of positive $\widetilde{\gamma}_N^{(2)}(j)$ contributes to the statistics, while only those $\widetilde{\gamma}_N^{(2)}(j)$ which are smaller than $\widetilde{\sigma}_{N,j,j}\sqrt{2\ln(\ln(N))/N}$ contribute to the critical values. This is why the coverage in the statement of Theorem 6 holds as a weak inequality. However, if for some j, $\gamma^{(2)}(j) = 0$, then the coverage is exactly $1-\alpha$, see Theorem 1 in Andrews and Guggenberger (2009). Hence, the simulated critical values provide a test with correct size for the limiting distribution under the least favorable case, and with unit asymptotic power.

5 Empirical Application: Neonatal Mortality in India

5.1 Data

The data we use is the second and the third-rounds of the District Level Household and Facility Survey (DLHS3 and DLHS2) from India.[13] DLHS3 (DLHS2) survey collected information from 720,320 (620,107) households residing in 612 (593) districts in 28 (29) states[14] and 6 union-territories (UTs) of India during the period 2007-08 (2002-04). These surveys focussed mainly on women and were designed to provide information on maternal and child health along with family planning and other reproductive health services. DLHS2 only included currently married women aged 15-44 but, DLHS3 included ever-married women aged 15-49 and never-married women aged 15-24. A multi-stage stratified sampling design that was representative at the district level was used. For our analysis we have combined both rounds of the surveys and have recoded the districts to match the boundaries that had been changed across the two surveys. This gave us 582 districts.

DLHS3 collected data on all pregnancies for each woman since 1st of January 2004. DLHS2 on the other hand collected information on all live births. The year and month of birth were recorded for all live births. For those children who had died by the time of the interview, year and month of death were also recorded. We convert this information to match the financial year (1st of April to 31st of March of the following year) in India as the conditional cash transfer program (CCT) of interest was administered at the beginning of a financial year.

[13] International Institute for Population Sciences (IIPS) was the nodal agency responsible for these surveys. Further details about the survey and relevant reports can be found at `http://www.rchiips.org/prch-3.html` and `http://www.rchiips.org/pdf/rch2/National_Report_RCH-II.pdf`.

[14] The state Nagaland was excluded in the third round.

5.2 The Conditional Cash Transfer Program-Janani Suraksha Yojana (JSY)

The National Rural Health Mission (NRHM) launched the program Janani Suraksha Yojana (JSY) in April 2005. This program replaced the National Maternity Benefit Scheme (NMBS) that had been available since August 1995. The objective of the JSY was to reduce maternal and neonatal mortality by promoting institutional delivery. NMBS was linked to the provision of better diet. However, JSY integrated cash assistance with antenatal care during pregnancy, followed by institutional care during delivery and immediate post-natal period (see Lingam and Kanchi (2013)). The scheme was rolled out from April 2005 with different districts adopting at different times.

The JSY program provided cash assistance to eligible pregnant women for delivery care (MOHFW, 2009). Initial financial assistance ranged between 500 to 1,000 Rupees (approx. 8 to 16 US Dollars) and has been modified over the years making it available to more women. The central government drew up the general guidelines for JSY in 2005. Whilst the adoption of JSY was compulsory for the whole of India, individual states were left with the authority to make minor alterations. The program was ultimately implemented by all the districts over time.

5.3 Sample and Variables

We do not have information on when and which districts implemented the program. We follow Lim et al. (2010) and Mazumdar et al. (2010) and create a treatment variable at the district level. The DLHS3 asked the mothers whether they had received financial assistance for delivery under the JSY scheme. Since the receipt of JSY could be correlated with unobserved mother specific characteristics in our model, we instead use this information to create a variable at the district level as follows. We define a district as having initiated the program in a particular year when the weighted[15] number of mothers who had received JSY among the mothers who gave birth in that district, exceeds a certain threshold for the first time. This district is defined as a 'treated' district from that period onwards. We experimented with different threshold. The main set of results are reported for the model using the 18% cutoff. The estimated effects were very similar across different thresholds.

There is a possibility that the States started the roll-out of the program in districts where the number of institutional deliveries were low and neonatal mortality was high. We therefore conduct our analysis using only the sample of babies born in the districts that were eventually treated during our observation period using the 18% cut-off. In addition, we have also extended the sample to include a few years prior to the program start to obtain enough deaths for the estimation of the baseline hazard. We use the birth and death information for babies born between April 2001 and December 2008 in these districts.

The object of interest is the deaths within the first 28 days after birth. Frequency distribution of the reported days of survival is presented in Table 1. We have 163,617 babies in our sample. Of these, 4,407 (2.69%) were recorded as having died before reaching

[15] As the DLHS is representative at the district level, appropriate weights to obtain summary statistics at the district level are provided in the dataset.

28 days implying a neonatal mortality rate of 441 babies per 1,000 live births. We make the following observations: (i) observed frequencies exhibit heaping at days which are multiples of 5; (ii) the number of reported deaths are quite small nearer the end of the time period. In order to model the baseline hazard non-parametrically, we also need enough exits over the observation period in the treated as well as the untreated districts. We therefore restrict our analysis to modelling the hazard during the first 18 days after birth. Hence our censoring point is 18 days instead of 28 days. The frequency distribution of survival information by treatment status is again provided in Table 1. 40,531 babies (24.8%) were born in the districts under treatment. The control group consists of 123,086 babies (75.2%). We also note, (i) the proportion of babies dying in each day is generally lower for babies born in treated districts compared to those born in untreated districts; (ii) the observed heaping at 5, 10, 15,..etc is still present in both samples.

As a preliminary to the estimation of formal models, it is informative to examine the non-parametric estimates of the unconditional hazard function distinguished by the treatment status. These are plotted in Figure 1. All babies born alive and survived the first 18 days are treated as censored observations and are also included in the risk set in the plot. The estimated hazard for those babies born in the treated districts generally lie below the hazard for the control group. The plots also show the heaping at durations which are multiples of 5 with a distinctive heap at 15 days and this is observed for both groups.

The model includes some control variables at the parental level as well as the child level. The parental level characteristics included are: (i) mother's age, mother's and father's education in years along with binary indicators for, caste and religion of the head of the household, whether the household lives in a rural area, and dummies for household standard of living is in the top or the middle third of the distribution.[16] The baby level characteristics included are binary indicators for sex, birth order and the year of birth. Summary statistics for these variables distinguished by treatment status are in Table 2. The average characteristics of the treated and the untreated samples are generally similar. The only differences are in those characteristics that have been improving over time. For example, the general level of schooling in India has been improving over time and hence the average years of schooling of parents in the treated sample are slightly higher given that the program was implemented recently.

5.4 Empirical Findings

We estimate the model using two different specifications. In the first one, we allow for heaping as outlined in Section 2, while the second one ignores this anomaly altogether. Since heaps in the data appear to be pronounced differently at different days (cf. Table 1), we allow for 'small' heaps at days 5 and 10, and for a 'big' heap at day 15 in the heaping specification. The former are associated with $\mathcal{D}^{\mathcal{H}_{-1}} = \{4, 9\}$ and $\mathcal{D}^{\mathcal{H}_{+1}} = \{6, 11\}$ together with the probabilities p_1 and p_3, while the 'big' heap is assumed to contain true durations from $\{13, 14\}$ and $\{16, 17\}$, respectively. The corresponding probabilities are p_1, p_2, and p_3, p_4, respectively. We set $\overline{k} = 12$ relying partially on information from the Program for

[16] The standard of living index was provided by the data people. See the DLHS2 (IIPS, 2006) and DLHS3 (IIPS, 2010) reports for further information about the construction of these indexes.

Appropriate Technology in Health (PATH) report (2012, p.20) on neonatal mortality in Uttar Pradesh, which suggests that the number of babies dying after 10 days after birth is relatively stable and not subject to large fluctuations.

Starting with the model allowing for heaping in Table 3 and the estimates of the probabilities p_1 to p_4, it is clear that in all four cases we reject the null hypothesis that the probability associated with the heaping process is equal to zero. The same results, albeit less pronounced, are obtained for corresponding tests on the probabilities $p_1^{(2)}$ to $p_4^{(2)}$, which are related to the heaping process after treatment and could differ from p_1 to p_4. Thus, allowing for heaping appears to be important within our setup, and, judging by the size of the estimated probabilities, treatment does not seem to have substantially altered the rounding patterns of individuals in the data (at least w.r.t. the specified heaps). Moreover, note that the estimated effects of the covariates in Table 3 are as expected.

Next, we turn to the maximum likelihood estimates of the $\gamma(\cdot)$ and $\gamma^{(2)}(\cdot)$ coefficients in Table 4. Theses parameters were estimated in exponential form ($\exp(\gamma(\cdot))$ and $\exp(\gamma^{(2)}(\cdot))$), which is shown in the first column of Table 4. Examining the size of the estimated coefficients, it is evident that, despite no real difference in their significance levels, $\exp(\gamma)$-coefficients are generally smaller in size in the model allowing for heaping than in the one without.

Turning to the effects of the JSY program and the second part of Table 4, we conduct the following tests in accordance with our theoretical results from Section 4 to gauge whether its introduction uniformly reduced mortality in the data across the first 18 days: using the IUP-GAMMA2 rule, we construct a battery of t-tests for each $\gamma^{(2)}$, e.g. $t_{\gamma_0^{(2)},N} = (0.947-1)/0.066 = -0.801$ for day 0 for the model with heaping, and compare each of these test statistics in turn with the one-sided 5% critical value from the normal distribution. Since it is obvious that we fail to reject some individual null hypotheses for $\gamma^{(2)} < 0$ as the exponential of the estimated coefficients is actually larger than one (which implies an increase in mortality), we cannot reject the union of the individual hypotheses either. Thus, we fail to reject the hypothesis that the introduction of the conditional cash transfer program reduced mortality uniformly across our observation period. On the other hand, we note that $\gamma^{(2)}(7)$ to $\gamma^{(2)}(10)$ as well as $\gamma^{(2)}(12)$ and $\gamma^{(2)}(16)$ are significantly less than one in the model allowing for heaping, which implies that we do reject $H2_0^{\gamma}$, the null that the cash transfer program had no effect on the γ coefficients over the period under examination. A similar conclusion can be drawn when examining the model without heaping, where a similar rejection pattern is observed. These effects are illustrated by Figures 2 and 3, which display the discrete hazard rates at $z_i = 0$ by treatment status for the model with and without heaping effects.

Summarizing the findings of this section, we note that our estimates suggest clear evidence of heaping in the data as the estimated 'heaping probabilities' were significantly different from zero. Moreover, the estimated coefficients in the model allowing for heaping were generally found to be smaller than the ones of the model without heaping. Finally, our test results did not indicate that the introduction of the JSY program reduced mortality uniformly over the first 18 days after birth. On the other hand, the program appears to have some effect after day 7, despite the low number of cases after the first week. Since our analysis was conducted using only those babies born in districts that were eventually treated, it remains to be established whether the actual effect of mothers receiving treatment exhibits a similar pattern, too. Drawing from the results of this paper, we conjecture that the program might not be targeting the mothers properly as districts that have implemented the program are not improving the survival chances of the babies substantially.

6 Conclusions

India has one of the largest neonatal mortality rates in the world. For this reason, the Indian Government launched a conditional cash-incentive program (JSY) to encourage institutional delivery in 2005. This paper studied the effect of the program on the neonatal mortality rate. Mortality is modeled using survival analysis, paying special attention to the substantial heaping present in the data. The main methodological contribution of the paper is the provision of a set of sufficient conditions for pointwise identification and consistent estimation of the baseline hazard in the joint presence of heaping and unobserved heterogeneity. Our identification strategy requires neither administrative data nor multiple measurements. It only requires the presence of a correctly reported duration and of some flat segments in the baseline hazard, which includes this correctly reported duration point. Information about the correctly reported duration and the flat segment can stem from different sources and does not need to come from a specific data set. The likelihood is constructed down-weighting the contribution of the heaped duration and over-weighting the contribution of the non heaped durations. This adjustment ensures consistent estimation of both heaping and baseline hazard parameters. We establish the asymptotic properties of the maximum likelihood estimator and provide simple procedure to test whether the policy had (uniformly) reduced mortality. Our empirical findings can be summarized as follows: first, heaping plays an important role in our data as the estimated probabilities associated with the heaping process were found to be significant before and after the introduction of the JSY program. Second, evidence for a uniform increase in survival probability of babies born in districts that were treated is rather scarce, despite a statistically significant increase after the first week. This casts some doubts about the overall efficacy of the JSY program in targeting mothers properly.

Appendix I

Proof of Proposition 1: In the following, suppose that $\overline{k} = \overline{\tau} - 2(\overline{r}+1)$ and $\overline{j}h^* = \overline{\tau} - \overline{r} - 1$, with $\overline{r} = 1$, so that $\overline{k} + 2 = \overline{j}h^* \in \mathcal{D}^{\mathcal{H}}$, and $\overline{\tau} = \overline{k} + 4$. The extension to $\overline{r} > 1$ will be outlined subsequently. Moreover, without loss of generality, assume that z_i is a scalar.

Define for any time period d

$$H_0(d) = \sum_{s=0}^{d} \exp(\gamma(s))$$

as the discrete cumulative baseline hazard.

First of all, notice that $\overline{k}$ is correctly observed by H(iii) and thus not in $\mathcal{D}^{\mathcal{H}}$, $\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}$, or $\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}$. This implies that:

$$\Pr\left(t_i = \overline{k} | z_i, \theta\right) = \Pr\left(\tau_i = \overline{k} | z_i, \theta\right).$$

Moreover, since time periods cannot belong to more than one heap (an immediate consequence of H(i) and the definition of the different sets), it must hold that:

$$\Pr\left(t_i \geq \overline{k} | z_i, \theta\right) = \Pr\left(\tau_i \geq \overline{k} | z_i, \theta\right).$$

Likewise, since individuals at $\overline{k} + 1$ only heap upwards, it also holds that:

$$\Pr\left(t_i \geq \overline{k} + 1 | z_i, \theta\right) = \Pr\left(\tau_i \geq \overline{k} + 1 | z_i, \theta\right).$$

For the case of correctly reported durations, we can proceed as in Heckman and Singer (p. 235, 1984). Given Assumption U,

$$\begin{aligned}
\Pr\left(\tau_i \geq \overline{k} + 1 | z_i, \theta\right) &= S_i\left(\overline{k} + 1 | z_i, \theta\right) \\
&= \int_0^\infty S_i\left(\overline{k} + 1 | z_i, v, \underline{\theta}\right) g(v; \sigma) dv \\
&= \int_0^\infty \exp\left(-v H_0(\overline{k}) \exp(z_i \beta)\right) g(v; \sigma) dv \\
&= \left(1 + \sigma\left(H_0(\overline{k}) \exp(z_i \beta)\right)\right)^{-\sigma^{-1}}.
\end{aligned}$$

Since the covariates are time invariant and independent of unobserved heterogeneity, set $z_i = 0$ to obtain

$$\begin{aligned}
S_i\left(\overline{k} + 1 | z_i = 0, \theta\right) &= \int_0^\infty \exp\left(-v H_0(\overline{k})\right) dG(v; \sigma) \\
&= \left(1 + \sigma\left(H_0(\overline{k})\right)\right)^{-\sigma^{-1}}.
\end{aligned}$$

Now $S_i\left(\overline{k} + 1 | z_i = 0, \theta\right)$ may be viewed as a composite of monotone functions, $A(H_0(\overline{k}))$, where:

$$A(H_0(\overline{k})) = \int_0^\infty \exp\left(-v H_0(\overline{k})\right) g(v; \sigma) dv.$$

To solve for $H_0(\overline{k})$, write $M = A(H_0(\overline{k}))$ and observe that $H_0(\overline{k}) = A^{-1}(M)$ is uniquely determined by strict monotonicity and continuity of A, which follows by U(iii) and the exponential form. Then, set $M = S_i\left(\overline{k}+1|z_i=0,\underline{\theta}\right)$ and deduce that:

$$H_0(\overline{k}) = A^{-1}(S_i\left(\overline{k}+1|z_i=0,\theta\right)).$$

Analogously,

$$H_0(\overline{k}-1) = A^{-1}(S_i\left(\overline{k}|z_i=0,\theta\right)),$$

and so $H_0(\overline{k}) - H_0(\overline{k}-1) = \exp(\gamma(\overline{k}))$, which identifies $\gamma(\overline{k})$. By assumption H(ii), this implies that also $H_0(\overline{k}+1) = H_0(\overline{k}+2)$ and $\gamma(\overline{k}+1) = \gamma(\overline{k}+2) = \gamma(\overline{k})$ are identified.

In the following, we will, without loss of generality, continue to set $z_i = 0$ for notational simplicity. Notice, however, that the argument carries through with $z_i \neq 0$ as β can be identified by standard arguments. Now, since the level of $\Pr\left(\tau_i = \overline{k}|z_i=0,\theta\right)$ is known and observed, and σ is identified by standard arguments, the probabilities

$$\Pr\left(\tau_i=\overline{k}+1|z_i=0,\theta\right) = \left(1+\sigma\left(H_0(\overline{k})\right)\right)^{-\sigma^{-1}} - \left(1+\sigma\left(H_0(\overline{k})\exp(\gamma(\overline{k}))\right)\right)^{-\sigma^{-1}}$$

and

$$\Pr\left(\tau_i=\overline{k}+2|z_i=0,\theta\right) = \left(1+\sigma\left(H_0(\overline{k})\exp(\gamma(\overline{k}))\right)\right)^{-\sigma^{-1}} - \left(1+\sigma\left(H_0(\overline{k})\exp(2\gamma(\overline{k}))\right)\right)^{-\sigma^{-1}}$$

are also known.

Moreover, to identify $\Pr\left(\tau_i=\overline{k}+2|z_i=0,\theta\right)$, notice that heaping in our setup is just a redistribution of probability masses between periods $\overline{k}+1$, $\overline{k}+2$, and $\overline{k}+3$. Thus, it holds that:

$$\begin{aligned}&\Pr\left(t_i=\overline{k}+1|z_i=0,\theta\right)+\Pr\left(t_i=\overline{k}+2|z_i=0,\theta\right)+\Pr\left(t_i=\overline{k}+3|z_i=0,\theta\right)\\ =&\Pr\left(\tau_i=\overline{k}+1|z_i=0,\theta\right)+\Pr\left(\tau_i=\overline{k}+2|z_i=0,\theta\right)+\Pr\left(\tau_i=\overline{k}+3|z_i=0,\theta\right)\end{aligned}$$

Hence, since the first two probabilities after the equality are known, we can identify $\Pr\big(\tau_i = \overline{k}+3|z_i=0,\theta\big)$ as:

$$\begin{aligned}&\Pr\left(\tau_i=\overline{k}+3|z_i=0,\theta\right)\\ =&\Pr\left(t_i=\overline{k}+1|z_i=0,\theta\right)+\Pr\left(t_i=\overline{k}+2|z_i=0,\theta\right)+\Pr\left(t_i=\overline{k}+3|z_i=0,\theta\right)\\ &-\Pr\left(\tau_i=\overline{k}+1|z_i=0,\theta\right)-\Pr\left(\tau_i=\overline{k}+2|z_i=0,\theta\right).\end{aligned}$$

In turn, by the same arguments as before, $\gamma(\overline{k}+3)$ can be identified from $H_0(\overline{k}+3) - H_0(\overline{k}+2)$.[17]

Finally, also p_1 and q_1 can be identified from:

$$\Pr\left(t_i=\overline{k}+1|z_i=0,\theta\right) = (1-p_1)\Pr\left(\tau_i=\overline{k}+1|z_i=0,\theta\right)$$

and

$$\Pr\left(t_i=\overline{k}+3|z_i=0,\theta\right) = (1-q_1)\Pr\left(\tau_i=\overline{k}+3|z_i=0,\theta\right).$$

[17]Note that $\Pr\left(t_i \geq \overline{k}+4|z_i,\theta\right) = \Pr\left(\tau_i \geq \overline{k}+4|z_i,\theta\right) = S_i\left(\overline{k}+4|z_i,\theta\right)$ is correctly observed as it is either the censoring point, is correctly observed, or belongs to $\mathcal{D}^{\mathcal{H}_{-l}}$ with $l=\overline{r}$.

Next examine the first heap for $j = 1$, i.e. h^*, and the corresponding times from $\mathcal{D}^{\mathcal{H}_{-1}}$ and $\mathcal{D}^{\mathcal{H}_{+1}}$. Since points from different heaps do not overlap by H(i), and periods prior to $h^* - 1$ are correctly observed, it holds that $\Pr(t_i \geq h^* - 1|z_i = 0, \theta) = \Pr(\tau_i \geq h^* - 1|z_i = 0, \theta) = S_i(h^* - 1|z_i = 0, \theta)$ and all γs up until $\gamma(h^* - 2)$ are identified.[18] Now,

$$
\begin{aligned}
& \Pr(t_i = h^* - 1|z_i = 0, \theta) \\
&= (1 - p_1) \Pr(\tau_i = h^* - 1|z_i = 0, \theta) \\
&= (1 - p_1) \left(\Pr(\tau_i \geq h^* - 1|z_i = 0, \theta) - \Pr(\tau_i \geq h^*|z_i = 0, \theta)\right) \\
&= (1 - p_1) \left(\int_0^\infty S_i(h^* - 1|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv - \int_0^\infty S_i(h^*|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv \right) \\
&= (1 - p_1) \left((1 + \sigma(H_0(h^* - 2)))^{-\sigma^{-1}} - (1 + \sigma(H_0(h^* - 1)))^{-\sigma^{-1}} \right),
\end{aligned}
$$

which uniquely identifies $H_0(h^* - 1)$, and so $\gamma(h^* - 1)$ since p_1, σ, all γs up until $\gamma(h^* - 2)$ have been already identified, and the above equation is strictly increasing and continuous in $H_0(h^* - 1)$.
Next, recalling

$$
\begin{aligned}
\Pr(t_i = h^* - 1|z_i = 0, \theta) &= (1 - p_1) \Pr(\tau_i = h^* - 1|z_i = 0, \theta), \\
\Pr(t_i = h^* + 1|z_i = 0, \theta) &= (1 - q_1) \Pr(\tau_i = h^* + 1|z_i = 0, \theta),
\end{aligned}
$$

and

$$
\begin{aligned}
& \Pr(t_i = h^*|z_i = 0, \theta) \\
&= p_1 \Pr(\tau_i = h^* - 1|z_i = 0, \theta) + \Pr(\tau_i = h^*|z_i = 0, \theta) + q_1 \Pr(\tau_i = h^* + 1|z_i = 0, \theta),
\end{aligned}
$$

it follows that

$$
\begin{aligned}
& \Pr(t_i = h^*|z_i = 0, \theta) - \frac{p_1}{1 - p_1} \Pr(t_i = h^* - 1|z_i = 0, \theta) \\
& -\frac{q_1}{1 - q_1} \Pr(t_i = h^* + 1|z_i = 0, \theta) \\
&= \Pr(\tau_i = h^*|z_i = 0, \theta) \\
&= \Pr(\tau_i \geq h^*|z_i = 0, \theta) - \Pr(\tau_i \geq h^* + 1|z_i = 0, \theta) \\
&= \left(\int_0^\infty S_i(h^*|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv - \int_0^\infty S_i(h^* + 1|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv \right) \\
&= \left((1 + \sigma(H_0(h^* - 1)))^{-\sigma^{-1}} - (1 + \sigma(H_0(h^*)))^{-\sigma^{-1}} \right)
\end{aligned}
$$

which uniquely identifies $\gamma(h^*)$, given that p_1, q_1, σ as well as $\gamma(s)$ for $s = 0, ..., h^* - 1$ have been already identified. As for $\gamma(h^* + 1)$,

$$
\begin{aligned}
& \Pr(t_i = h^* + 1|z_i = 0, \theta) \\
&= (1 - q_1) \left(\Pr(\tau_i \geq h^* + 1|z_i = 0, \theta) - \Pr(\tau_i \geq h^* + 2|z_i = 0, \theta)\right) \\
&= (1 - q_1) \left(\int_0^\infty S_i(h^* + 1|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv - \int_0^\infty S_i(h^* + 2|z_i = 0, v, \underline{\theta}) g(v; \sigma) dv \right) \\
&= (1 - q_1) \left((1 + \sigma(H_0(h^*)))^{-\sigma^{-1}} - (1 + \sigma(H_0(h^* + 1)))^{-\sigma^{-1}} \right),
\end{aligned}
$$

[18] If $h^* = 1$, $S_i(h^* - 1|z_i, \theta) = 1$ by definition.

which uniquely identifies $\gamma(h^*+1)$. The remaining heaps follow analogously.

We will now consider the extension to $\overline{r} > 1$: $\gamma(\overline{k})$ can be identified as before and thus we can construct $\Pr\left(\tau_i = \overline{k}+1|z_i=0,\theta\right) = \ldots = \Pr\left(\tau_i = \overline{k}+\overline{r}+1|z_i=0,\theta\right) = \Pr\left(\tau_i = \overline{k}|z_i=0,\theta\right)$. Next, observe that:

$$\begin{aligned}
&\sum_{l=1}^{\overline{r}} \Pr\left(t_i = \overline{k}+l|z_i=0,\theta\right) + \Pr\left(t_i = \overline{k}+\overline{r}+1|z_i=0,\theta\right) \\
&+ \sum_{l=1}^{\overline{r}} \Pr\left(t_i = \overline{k}+\overline{r}+1+l|z_i=0,\theta\right) \\
=&(\overline{r}+1)\Pr\left(\tau_i = \overline{k}|z_i=0,\theta\right) + \overline{r}\Pr\left(\tau_i = \overline{k}+\overline{r}+2|z_i=0,\theta\right),
\end{aligned}$$

where we used the fact that γ is constant after $\overline{k}+\overline{r}+2$. Thus, since $\overline{r}$ is known,

$$\begin{aligned}
&\Pr\left(\tau_i = \overline{k}+\overline{r}+2|z_i=0,\theta\right) \\
=&\frac{1}{\overline{r}}\left[\sum_{l=1}^{\overline{r}} \Pr\left(t_i = \overline{k}+l|z_i=0,\theta\right) + \Pr\left(t_i = \overline{k}+\overline{r}+1|z_i=0,\theta\right)\right. \\
&\left. + \sum_{l=1}^{\overline{r}} \Pr\left(t_i = \overline{k}+\overline{r}+1+l|z_i=0,\theta\right) - (\overline{r}+1)\Pr\left(\tau_i = \overline{k}|z_i=0,\theta\right)\right]
\end{aligned}$$

is identified. Hence, for each $l = 1,\ldots,\overline{r}$, we can now retrieve the probabilities from:

$$\Pr\left(t_i = \overline{k}+l|z_i=0,\theta\right) = (1-p_l)\Pr\left(\tau_i = \overline{k}+l|z_i=0,\theta\right)$$

and

$$\Pr\left(t_i = \overline{k}+\overline{r}+1+l|z_i=0,\theta\right) = (1-q_l)\Pr\left(\tau_i = \overline{k}+\overline{r}+1+l|z_i=0,\theta\right).$$

as before. Then, examining the first heap again, note that each γ prior to $\gamma(h^*-l)$ can now be uniquely identified from

$$\begin{aligned}
&\Pr\left(t_i = h^*-l|z_i=0,\theta\right) \\
&= \; (1-p_l)\Pr\left(\tau_i = h^*-l|z_i=0,\theta\right),
\end{aligned}$$

$\gamma(h^*-l)$ can be uniquely identified from

$$\begin{aligned}
&\Pr\left(t_i = h^*|z_i=0,\theta\right) - \sum_{l=1}^{\overline{r}} \frac{p_l}{1-p_l}\Pr\left(t_i = h^*-l|z_i=0,\theta\right) \\
&- \sum_{l=1}^{\overline{r}} \frac{q_l}{1-q_l}\Pr\left(t_i = h^*+l|z_i=0,\theta\right) \\
=\;& \Pr\left(\tau_i = h^*|z_i=0,\theta\right),
\end{aligned}$$

and so on, by the same argument used for $\overline{r}=1$.

Proof of Theorem 2:
(i) Given Assumption D, by the uniform law of large number for identically and independently distributed observations,

$$\sup_{\theta\in\Theta} \left| \left(l_N(\theta) - \mathrm{E}\left(l_N(\theta) \right) \right) / N \right| = o_p(1)$$

and recalling that the argmax is a continuous function,

$$\arg\max_{\theta\in\Theta} l_N(\theta) - \arg\max_{\theta\in\Theta} \mathrm{E}\left(l_N(\theta) \right) = o_p(1).$$

As $\theta^{\dagger} = \arg\max_{\theta\in\Theta} E\left(l_N(\theta) \right)$, and $\theta^{\dagger}$ is unique, because of the unique identifiability established in Proposition 1, the statement in (i) follows.
(ii) The statement follows from Theorem 3(a)-(b) in Andrews (1999), hereafter A99, once we show that his Assumption 2-6 hold. Note that, given Assumption U,

$$\begin{aligned} \int \Pr\left(t_i > t | z_i, v, \underline{\theta} \right) g(v;\sigma) dv &= \int S_i\left(t | z_i, v, \underline{\theta} \right) g(v;\sigma) dv \\ &= \left(1 + \sigma \left(\sum_{s=0}^{t-1} \exp\left(z_i' \beta + \gamma(s) \right) \right) \right)^{-\sigma^{-1}}, \end{aligned}$$

and from the definition of $\phi_i(\cdot)$ in (i)-(v), it is immediate to see that $l_N(\theta)$ has well defined left and right derivatives for $\theta \in \Psi^{+}$, with $\Psi^{+} = \Psi \cap C\left(\theta^{\dagger}, \varepsilon\right)$, with $C\left(\theta^{\dagger}, \varepsilon\right)$ denoting an open cube of radius ε around $\theta^{\dagger}$. Thus $l_N(\theta)$ has the following quadratic expansion

$$l_N(\theta) - l_N(\theta^{\dagger}) = \nabla_{\theta} l_N(\theta^{\dagger}) \left(\theta - \theta^{\dagger} \right) + \frac{1}{2} \left(\theta - \theta^{\dagger} \right)' \nabla^2_{\theta\theta} l_N(\theta^{\dagger}) \left(\theta - \theta^{\dagger} \right) + R_N(\theta),$$

with $R_N(\theta) = O_p\left(N^{-3/2} \right)$, because of the existence of third order partial left and right derivatives. This ensures that Assumption 2^{*} in A99 is satisfied, which in turn implies Assumption 2 in A99 holds too. By the central limit theorem for *iid* random variables, and given the information matrix equality, $N^{-1/2} I_N^{-1} \nabla l_N(\theta^{\dagger}) \xrightarrow{d} N\left(0, \mathcal{I}^{\dagger -1}\right)$. This ensures that Assumption 3 in A99 holds. Given the consistency established in part (i), Assumption 4 in A99 follows immediately from his Assumptions A2* and A3. Given Assumption H(iv), the boundary issue which may arise is when some p_l and/or q_l for $l = 1, ..., \overline{r}$ are zero. Hence, $\left(\Theta - \theta^{\dagger} \right)$ is locally equal to Ψ which is a convex cone in $R^{p_\beta + \overline{k} + 2 + 2\overline{r}}$, and Assumption 5 and 6 in A99 hold.
(iii) In this case $\theta^{\dagger}$ is not on the boundary, and so

$$\widehat{\psi} = \inf_{\psi\in\Psi} \left((\psi - G)' \mathcal{I}^{\dagger} (\psi - G) \right) = G$$

and $G \sim N\left(0, \mathcal{I}^{\dagger -1}\right)$.

Proof of Proposition 3:
Given Assumption D', the statements in Theorem 2 hold with $\widehat{\theta}_N$ replaced by $\widetilde{\vartheta}_N$, and $\theta^{\dagger}$ replaced by $\vartheta^{\ddagger}$, thus for each $j = 1, ..., \overline{r}$,

$$\widehat{p}_{j,N} \xrightarrow{d} \max\left\{0, G_j\right\}, \ G_j \sim N(0, \mathcal{I}^{-1}_{p_j p_j}),$$

and as, given Assumption D', $\widehat{\mathcal{I}}_N - \mathcal{I} = o_p(1)$, $t_{p_j,N} \overset{d}{\to} N(0,1)$, and so rejecting whenever $PV_{p,j,N} < \alpha$ ensures a level α test for $H_{p,0}^{(j)}$. It then follows from the same argument in the proof of Proposition 5.3.1 in Silvapulle and Sen (2005), that the overall size of the test is at most α. Finally, the power is one as the probability of failing to reject each single $H_{p,0}^{(j)}$ or $H_{q,0}^{(j)}$ is asymptotically one.

Proof of Proposition 4
By the same argument as in Proposition 3.

Proof of Theorem 5: By the law of large numbers and the central limit for independent non-identical series,

$$\sqrt{N}\left(\frac{\widetilde{\gamma}_N^{(2)}(0) - \gamma^{(2)}(0)}{\widetilde{\sigma}_{N,0,0}}, ..., \frac{\widetilde{\gamma}_N^{(2)}(\overline{k}) - \gamma^{(2)}(\overline{k})}{\widetilde{\sigma}_{N,\overline{k},\overline{k}}}, ...,\right) \overset{d}{\to} N\left(0, \Omega\right),$$

with $\Omega = D^{-1/2}\left(R\Sigma R'\right)D^{-1/2}$ as defined in the statement of the Theorem. By noting that,

$$S_N^- = \sum_{j=0}^{\overline{k}+1}\left(\sqrt{N}\frac{\widetilde{\gamma}_N^{(2)}(j) - \gamma^{(2)}(j)}{\widetilde{\sigma}_{N,j,j}} + \sqrt{N}\frac{\gamma^{(2)}(j)}{\widetilde{\sigma}_{N,j,j}}\right)_-^2,$$

the statement under $H2_0^\gamma$ follows by the continuous mapping theorem, as S_N^- satisfies Assumption 1-3 in Andrews and Guggenberger (2009). Under $H2_A^\gamma$ there is some j such that $\gamma^{(2)}(j) < 0$ and then the statistic diverges at rate $\sqrt{N}$.

Proof of Theorem 6:

By the law of the iterated logarithm, as $N \to \infty$, with probability approaching one, for $j = 0, ...\overline{k}+1$, $\left(\frac{N}{2\ln\ln N}\right)^{1/2}\frac{\widetilde{\gamma}_N^{(2)}(j)}{\widetilde{\sigma}_{N,j,j}} \leq 1$ if $\gamma^{(2)}(j) = 0$ or if $\gamma^{(2)}(j) < 0$, while $\left(\frac{N}{2\ln\ln N}\right)^{1/2}\frac{\widetilde{\gamma}_N^{(2)}(j)}{\widetilde{\sigma}_{N,j,j}} > 1$ if $\gamma^{(2)}(j) > 0$. Hence, when $H2_0^\gamma$ is true, as N gets large only those $\widetilde{\gamma}_N^{(2)}(j)$ associated with $\gamma^{(2)}(j) = 0$ contribute to the simulated limiting distribution, and in the meantime, the probabillity of eliminating a non-slack ("too" positive) $\widetilde{\gamma}_N^{(2)}(j)$ approaches zero. This ensures that the statement in (4) holds, and holds as strict equality if for some j, $\gamma^{(2)}(j) = 0$. The statement in (5) follows immediately, as for $b = 1, ..., B$ $S_N^{-*(b)}$ has a well defined limiting distribution under both hypotheses, while S_N^- diverges to infinity under the alternative.

Appendix II

Table 1: Neonatal Mortality − Deaths by Number of Days of Survival

				Births Untreated Dist.[1]		Births Treated Dist.[1]	
Days	Freq.	Percent	Cum. Percent	Freq.	Percent	Freq.	Percent
0	1,065	0.65	0.65	832	0.68	233	0.57
1	1,097	0.67	1.32	853	0.69	244	0.6
2	358	0.22	1.54	261	0.21	97	0.24
3	426	0.26	1.8	325	0.26	101	0.25
4	187	0.11	1.91	144	0.12	43	0.11
5	201	0.12	2.04	154	0.13	47	0.12
6	103	0.06	2.1	81	0.07	22	0.05
7	130	0.08	2.18	104	0.08	26	0.06
8	160	0.10	2.28	133	0.11	27	0.07
9	45	0.03	2.31	35	0.03	10	0.02
10	106	0.06	2.37	85	0.02	21	0.05
11	37	0.02	2.39	29	0.02	8	0.02
12	41	0.03	2.42	34	0.03	7	0.02
13	17	0.01	2.43	14	0.01	3	0.01
14	28	0.02	2.45	20	0.02	8	0.02
15	180	0.11	2.56	150	0.12	30	0.07
16	21	0.01	2.57	18	0.01	3	0.01
17	13	0.01	2.58	8	0.01	5	0.01
18	19	0.01	2.59				
19	8	0	2.59				
20	71	0.04	2.64				
21	28	0.02	2.65				
22	18	0.01	2.66				
23	5	0	2.67				
24	5	0	2.67				
25	26	0.02	2.69				
26	7	0	2.69				
27	5	0	2.69				
Cens. Obs.	159,210	97.31	100	119,806	97.38	39,596	97.70
Total	163,617	100		123,086	100	40,531	100

[1] The treatment status is based on whether at least 18% of the women who gave birth in a particular financial year said that they had received cash under the program JSY.

Figure 1: Unconditional Hazards

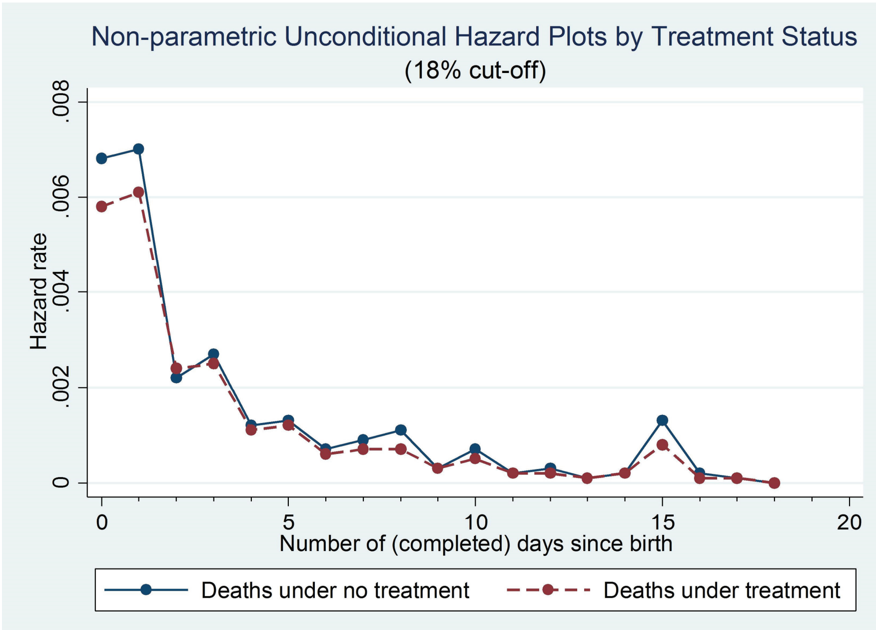


Table 2: Summary Statistics of Covariates

			Untreated[1]		Treated[1]	
Number of Obs Children	163,617		123,086		40,531	
Number of Obs Mothers	127,637		96154		31,377	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Parental Characteristics						
Mother's Age (years)	25.62	5.21	25.90	5.21	24.74	4.97
Mother's Schooling (years)	4.52	4.79	4.32	4.78	5.11	4.79
Father's Schooling (years)	6.49	4.98	6.43	5.02	6.70	4.85
Caste: Base − Other Backward						
Schedule Caste	0.18		0.18		0.18	
Schedule Tribe	0.18		0.18		0.18	
Privileged Caste	0.24		0.24		0.23	
Religion: Base − Hindu						
Muslim	0.10		0.10		0.10	
Other	0.08		0.08		0.08	
Living Std.: Base − Bottom Third						
Milddle third	0.33		0.32		0.36	
Top third	0.22		0.20		0.25	
Rural Household	0.78		0.77		0.81	
Child Characteristic						
Girl	0.48		0.48		0.48	
Birth Order: Base − First Born						
Bord=2	0.28		0.28		0.30	
Bord=3	0.16		0.16		0.14	
Bord 4 or more	0.20		0.21		0.16	
Birth Year: Base − 2001						
2002	0.11		0.15		0.00	
2003	0.12		0.15		0.00	
2004	0.14		0.19		0.00	
2005	0.15		0.18		0.04	
2006	0.15		0.09		0.35	
2007	0.15		0.01		0.58	
2008	0.01		0.00		0.04	

[1] We define babies as 'treated' if they are born in a district where at least 18% of the women who had given birth said they had received cash under the program JSY.

Table 3: Estimated Effects of Covariates & 'Heaping' Probabilities

Model with Heaping[1]		
	Coeff.Est.	Bootstrapped S.E.[2]
Before Treatment		
p_1	0.539	0.068
p_2	0.466	0.052
p_3	0.373	0.063
p_4	0.498	0.058
After Treatment		
$p_1^{(2)}$	0.412	0.106
$p_2^{(2)}$	0.558	0.052
$p_3^{(2)}$	0.468	0.078
$p_4^{(2)}$	0.325	0.100
Covariates[3]		
Mother's age	– 0.108	0.011
Mother's schooling	– 0.016	0.013
Father's schooling	– 0.056	0.016
Girl	– 0.159	0.045
Birth Order 2	– 0.226	0.049
Birth Order 3	– 0.140	0.054
Birth Order 4 or higher	0.110	0.080
2002	– 0.033	0.064
2003	– 0.002	0.054
2004	– 0.203	0.043
2005	– 0.153	0.055
2006	– 0.150	0.043
2007	– 0.178	0.049
2008	– 0.228	0.085
Scheduled caste	– 0.098	0.056
Scheduled tribe	– 0.280	0.048
Privileged caste	– 0.120	0.043
Muslim	– 0.255	0.049
Other religion	– 0.256	0.050
Middle third std. living	– 0.069	0.039
Top third std. living	– 0.014	0.067
Rural household	– 0.156	0.046

[1] Model allows for small heaps at days 5 and 10 with associated probabilities p_1 and p_3, and a large heap at day 15 with associated probabilities p_1, p_2, p_3, and p_4. $\bar{k}$ was set to $\bar{k} = 12$.

[2] Bootstrapped standard errors with 100 replications (see Appendix III for details).

[3] For testing purposes, the effects of the covariates on the hazard before and after the introduction of JSY were assumed to be the same.

Table 4: Maximum Likelihood Estimates

	Model with Heaping[1]		Model without Heaping	
	Coeff.Est.	Bootstrapped S.E.[2]	Coeff.Est.	Bootstrapped S.E.[2]
$\exp(\gamma)$ by day	Before Treatment			
0	0.484	0.048	0.573	0.044
1	0.492	0.049	0.586	0.046
2	0.292	0.026	0.340	0.024
3	0.306	0.031	0.369	0.027
4	0.243	0.026	0.235	0.019
5	0.176	0.022	0.254	0.018
6	0.160	0.021	0.186	0.015
7	0.170	0.017	0.206	0.017
8	0.195	0.021	0.236	0.019
9	0.115	0.024	0.116	0.022
10	0.145	0.021	0.189	0.016
11	0.102	0.023	0.109	0.025
12	0.115	0.029	0.122	0.025
13	0.115	0.029	0.204	0.036
14	0.115	0.029	0.096	0.032
15	0.115	0.029	0.250	0.017
16	0.073	0.035	0.085	0.032
17	0.073	0.035	0.064	0.036
$\exp(\gamma^{(2)})$ by day	Change after Treatment[3]			
0	0.947	0.066	0.986	0.074
1	0.983	0.067	1.107	0.067
2	1.001	0.072	1.098	0.068
3	0.928	0.080	1.069	0.083
4	0.867	0.088	0.975	0.107
5	0.927	0.098	0.999	0.100
6	0.860	0.087	0.880	0.088
7	0.819	0.093	0.854	0.082
8	0.752	0.075	0.770	0.077
9	0.829	0.084	0.798	0.066
10	0.723	0.074	0.837	0.085
11	0.891	0.070	0.947	0.073
12	0.822	0.049	0.771	0.067
13	0.822	0.049	0.754	0.062
14	0.822	0.049	0.815	0.080
15	0.822	0.049	0.785	0.059
16	0.844	0.062	0.720	0.059
17	0.844	0.062	0.763	0.071
σ[4]	0.710	0.054	0.781	0.057

[1] Model allows for small heaps at days 5 and 10 with associated probabilities p_1 and p_3, and a large heap at day 15 with associated probabilities p_1, p_2, p_3, and p_4. $\overline{k}$ was set to $\overline{k} = 12$.
[2] Bootstrapped standard errors with 100 replications (see Appendix III for details).
[3] $\gamma^{(2)} = 1$ implies that treatment had no effect.
[4] σ is the inverse of the variance of the gamma distributed unobserved heterogeneity.

Figure 2: Estimated Hazard without Heaping

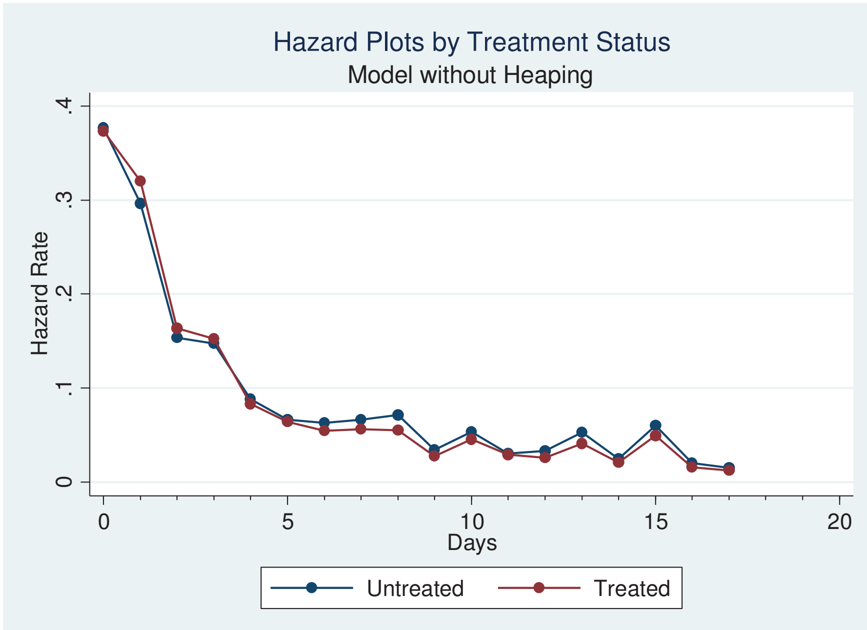
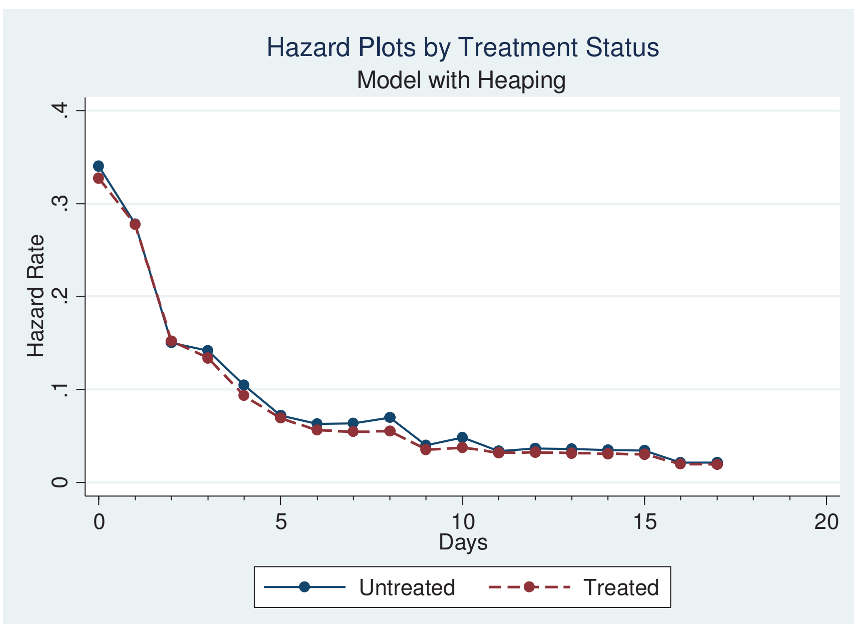


Figure 3: Estimated Hazard with Heaping



Appendix III

The standard errors in the illustration of Section 5 have been constructed using the bootstrap method. However, due to the possibility of one or more parameters lying on the boundary of the parameter space, which invalidates the first order validity of the naive bootstrap (see Andrews, 2000), we follow Section 6.4 in Andrews (1999) and construct standard errors based on subsampling. More precisely, as we resample with replacement we are implementing m out of n ($moon$) bootstrap.

Let $\widetilde{l}_i(\vartheta)$ be the contribution of baby i-th to the likelihood $\widetilde{l}_N(\vartheta)$. Let I_j, $j = 1, ..., M$, be M independent draws from a discrete uniform on $[1, N]$. We then make M draws, with replacement from $\left(\widetilde{l}_1(\vartheta), ..., \widetilde{l}_N(\vartheta)\right)$, to get $\left(\widetilde{l}_{I_1}(\vartheta), ..., \widetilde{l}_{I_M}(\vartheta)\right) = \left(\widetilde{l}_1^*(\vartheta), ..., \widetilde{l}_M^*(\vartheta)\right)$. Note that for M sufficiently large, the proportion of draws before and after treatment matches the sample proportion, and this ensure the validity of the bootstrap even in the presence of a possible structural break due to the treatment effect. Indeed, Goncalves and White (2004) suggest to resample the likelihood instead of directly resampling the observations, in order to deal with possible heterogeneity. Let

$$\widetilde{\vartheta}_M^* = \arg\max_{\vartheta \in \widetilde{\Theta}} \sum_{j=1}^{M} \widetilde{l}_i^*(\vartheta)$$

and let $\left(\widetilde{\vartheta}_M^{*(1)}, ..., \widetilde{\vartheta}_M^{*(B)}\right)$ denotes the bootrap estimator at replication $1, ..., B$. Now, the estimator of the bootstrap variance-covariance matrix reads as

$$\widehat{V}_{M,B}^* = \frac{M}{B} \sum_{j=1}^{M} \left(\left(\widetilde{\vartheta}_M^{*(j)} - \frac{1}{M} \sum_{j=1}^{M} \widetilde{\vartheta}_M^{*(j)} \right) \left(\widetilde{\vartheta}_M^{*(j)} - \frac{1}{M} \sum_{j=1}^{M} \widetilde{\vartheta}_M^{*(j)} \right)' \right).$$

Given Assumptions H,U,C and D', we can show that the conditions in Theorem 1 in Goncalves and White (2005) are satisfied.[19] It then follows that as $M/N \to 0$, $N, M, B \to \infty$

$$\widehat{V}_{M,B}^* - \Sigma = o_P(1) + o_{p^*}(1),$$

where $\Sigma = \lim_{N\to\infty} \text{var}\left(\sqrt{N}\left(\widetilde{\vartheta}_N - \vartheta^\ddagger\right)\right)$, and $o_{p^*}(1)$ denotes a term converging to zero according to the bootstrap probability law, as $M, B \to \infty$. Hence, the standard error for the elements of $\widetilde{\vartheta}_N$ can be obtained using the square root of the diagonal element of $\frac{1}{M}\widehat{V}_{M,B}^*$.

It remains to select M and B. In our set-up we need to choose M rather large relative to N. This is to ensure we have enough "exits" for each duration. In practice M is roughly equal to $0.8N$, which may violate the condition $M/N \to \infty$. However, for smaller value of the ratio M/N we do not have enough exits and so we would violate Assumption H(iv), which is necessary for identification. Finally, given the highly nonlinearity of our model, and the large number of parameters, we have to set $B = 100$. [20]

[19] This can be establisheded using the same argument as in Appendix B in the Supplementary Material of Corradi, Distaso and Mele (2013).

[20] Roughly speaking each bootstrap iteration takes about 7 minutes, and thus we need to limit the number of replications. As a robustness check, we have tried much larger values of B, $B = 300, 500$ for a simplified version of the model, with a much smaller number of covariates. Our findings are quite robust to the choice of B.

References

Abbring, J.H. and G. Van Den Berg (2007). The Unobserved Heterogeneity Distribution in Duration Analysis. *Biometrica*, 94, 87-99.

Abrevaya, J. and J. Hausman (1999). Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells. *Annales d'Economie et de Statistique,* 55/56, 243-275.

Andrews, D.W.K. (1999). Estimation When a Parameter is on the Boundary. *Econometrica,* 67, 1341-1383.

Andrews, D.W.K. (2000). Inconsistency of the Bootstrap when a Parameters is on the Boundary of the Parameter Space. *Econometrica,* 68, 399-405.

Andrews, D.W.K. and P. Guggenberger (2009). Validity of Subsampling and "Plug-In Asymptotic" Inference for Parameters Defined by Moment Inequalities. *Econometric Theory*, 25, 669-709.

Andrews, D.W.K. and G. Soares (2010). Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica*, 78, 119-158.

Bierens, H.J. (2008). Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results. *Econometric Theory,* 24, 749-794.

Burda, M., M. Harding, and J. Hausman (2014). A Bayesian Semiparametric Competing Risk Model With Unobserved Heterogeneity. *Journal of Applied Econometrics*, preprint.

Chernozhukov, V., H. Hong and E. Tamer (2007). Estimation and Confidence Regions for Parameters Sets in Econometric Models. *Econometrica*, 75, 1243-1284.

Corradi, V. and W. Distaso (2011). Multiple Forecast Model Evaluation, *Oxford Handbook of Forecasting*, edited by M.P. Clements and D.F, Hendry, Oxford University Press.

Corradi, V., W. Distaso and A. Mele (2013). Macroeconomic Determinants of Stock Volatility and Volatility Premiums. *Journal of Monetary Economics,* 60, 203-220 (Supplementary Material).

Cox, D.R. and D. Oakes (1984). *Analysis of Survival Data*, Boca Raton, Chapman & Hall.

Elbers, C. and G. Ridder (1982). True and Spurious Duration Dependence: The Identifiability of the Mixed Proportional Hazard Model. *Review of Economic Studies.* 49, 403-411.

Goncalves, S. and H. White (2004). Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models. *Journal of Econometrics.* 119, 199-220.

Goncalves, S. and H. White (2005). Bootstrap Standard Error Estimates for Linear Regression. *Journal of the American Statistical Association.* 471, 970-979.

Ham, J.C., X. Li and L. Shore-Sheppard (2014). Seam Bias, Multiple-State, Multiple-Spell Duration Models and the Employment Dynamics of Disadvantaged Women. University of Maryland, Working Paper.

Han, A. and J. Hausman (1990). Flexible Parametric Estimation of Duration and Competing Risk Models, *Journal of Applied Econometrics*, 5, 1-28.

Hausman, J. and T. Woutersen (2013). Estimating a Semiparametric Duration Model without Specifying Heterogeneity. *Journal of Econometrics,* forthcoming.

Heitjan, D.F. and Rubin, D.B. (1990). Inference From Coarse Data Via Multiple Imputation With Application to Age Heaping, *Journal of the American Statistical Association*, 85, 304-314.

Heckman, J. and B. Singer (1984). The Identifiability of the Proportional Hazard Model. *Review of Economic Studies,* 51, 231-241.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

International Institute for Population Sciences (2006). *District Level Household Survey (DLHS-2), 2002-04: India*. Mumbai: IIPS.

International Institute for Population Sciences (2010). *District Level Household and Facility Survey (DLHS-3), 2007-08: India.* Mumbai: IIPS.

Lim, S., Dandona, L., Hoisington, J., James, S., Hogan, M., and Gakidou, E. (2010). India's Janani Suraksha Yojana, A Conditional Cash Transfer Programme to Increase Births in Health Facilities: An Impact Evaluation. *The Lancet*, 375(9730), 2009–2023.

Lingam, L. and Kanchi, A. (2013). *Women's Work, Maternity and Public Policy in India*, TATA Institute of Social Sciences, Hyderabad, India.

Mazumdar, S., Mills, A., and Powell-Jackson, T. (2011). Financial Incentives in Health: New Evidence from India's Janani Suraksha Yojana. Unpublished Manuscript.

Mazumdar, S., Mills, A., and Powell-Jackson, T. (2011). Financial Incentives in Health: New Evidence from India's Janani Suraksha Yojana. Unpublished Manuscript.

Meyer, B. D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, 58, 757-782.

Ministry of Health and Family Welfare (MOHFW) (2009). *Four Years of NRHM, 2005-2009: Making a Difference Everywhere*. MOHFW, India.

Program for Appropriate Technology in Health (PATH) (2012). *Understanding Neonatal Mortality in Uttar Pradesh, India: Identifying Underlying Causes of Death Through Verbal Autopsy* . PATH, Seattle (WA).

Petoussi, K., R.D. Gill and C. Zeelenberg (1997). Statistical Analysis of Heaped Duration Data. Vrije University. Working Paper.

Ridder, G., T.M. Woutersen (2003). The Singularity of the Information Matrix of the Mixed Proportional Hazard Model. *Econometrica*, 71, 15791589.

Romano, J.P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73, 1237-1282.

Roy, S.S. and Others (2013). Improved Neonatal Survival After Participatory Learning and Action With Womens Groups: A Prospective Study in Rural Eastern India. *Bulletin World Health Organization*, 91, 426433B.

Silvapulle, M.J. and P.K. Sen (2005). *Constrained Statistical Inference*, New York, Wiley.

Szlydlowski, A. (2013). Endogenous Censoring in the Mixed Proportional Hazard Model with an Application to Optimal Unemployment Insurance. Working Paper, Northwestern University.

Torelli, N. and U. Trivellato (1993). Modelling Inaccuracies in Job-Search Duration Data. *Journal of Econometrics,* 59, 187-211.

White, H. (2000). A Reality Check For Data Snooping. *Econometrica*, 68, 1097-1127.

Wolff, J. and T. Augustin (2000). Heaping and Its Consequences for Duration Analysis. Working Paper 203, Ludwig Maximilians University.