Modeling Heaped Duration Data: An Application to Neonatal Mortality*

Wiji Arulampalam† Warwick University
Valentina Corradi ‡ Surrey University
Daniel Gutknecht§ Mannheim University

April 12, 2016

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 substantial heaping, a form of measurement error, 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 (discretized) baseline hazard accounting for heaping and unobserved heterogeneity. Our identification strategy requires neither administrative data nor multiple measurements, but a correctly reported duration point and the presence of some flat segment(s) in the baseline hazard. We establish the asymptotic properties of the maximum likelihood estimator and derive a set of specification tests that allow, among other things, to test for the presence of heaping and to compare different heaping mechanisms. Our empirical findings do not suggest a significant reduction of mortality in treated districts. However, they do indicate that accounting for heaping matters for the estimation of the hazard parameters.

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

JEL Classification: C12, C21, C24, C41.

*We are grateful to the editors Tom Wansbeek and Yingyao Hu and two anonymous referees as well as Aureo de Paula, Giovanni Forchini, Federico Martellosio, and Mervin Silvapulle for many useful comments and suggestions that helped to greatly improve the paper. We would also like to thank Marco Alfano and Artemisa Flores for help with the data preparation. Comments received from seminar participants at Oxford University, the 6th Italian Conference on Econometrics and Empirical Economics (ICEEE) in Salerno, Groningen University, CEMFI, ENSAI, Mannheim University, Munich University, and Leicester University are gratefully acknowledged.

†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 Mannheim, L7, 3-5, 68131 Mannheim, Germany. Email: `Daniel.Gutknecht@gmx.de`

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 each year (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 each year. In order to tackle this huge problem, the Indian Government introduced a conditional cash-incentive (*Janani Suraksha Yojana* - JSY) 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 the above Indian cash-incentive program on neonatal mortality using district-level household survey data. We focus on the neonatal period, since the effects of the program are 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 apparent heaping at $5, 10, 15, \dots$ days (durations which are multiple of five days). These heaping effects in the data are likely to be due to measurement error and lead, if neglected, to inconsistency in the estimation of the underlying hazard function (e.g., Torelli and Trivellato, 1993, and Wolff and Augustin, 2003).

In addressing heaping effects due to misreporting, this paper makes several methodological contributions in the modeling of discrete time duration data when the data is characterized by abnormal concentrations at certain duration points. First, we provide a set of sufficient conditions for the pointwise identification and consistent estimation of the (discretized) baseline hazard and other parameters of a proportional hazard model 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. Second, we derive various specification tests to test (i) for the presence as well as (ii) for different patterns of heaping effects in our model, and (iii) to assess the effects of policy changes on the baseline hazard.[2] These tests provide the applied researcher with a set of tools that enable her to verify the validity of different model specifications.

Despite the prevalence of heaping in survey data, the econometric literature on identification and estimation of duration models under heaping is rather limited. In the presence of misreported durations, Abrevaya and Hausman (1999) derive a set of sufficient conditions under which the monotone rank estimator remains a consistent estimator for the covariate parameters of an accelerated failure time and proportional hazard model. Torelli and Trivellato (1993) and Augustin and Wolff (2004) derive procedures that allow for different forms of heaping, but assume a parametric specification for the hazard function, which is not suitable if the ultimate goal of an analysis is inference on a policy change that might not have affected the entire hazard function. Petoussis, Gill and Zeelenberg (1997) treat heaped durations as missing values and use the Expectation-Maximization (EM) algorithm

[1] The neonatal period is defined as the first 28 days after birth.

[2] We also show that a test on parameter equality between a model that takes account of heaping and a model that neglects heaping is equivalent to the test on the presence of heaping effects (see Section S2.2 of the supplementary material).

to estimate a proportional hazard model. Finally, Heitjan and Rubin (1990) suggest an EM-based multiple imputation method for inference in the presence of heaped data. The authors do, however, not cover the duration case. Moreover, note that none of the above papers deals with identification of the baseline hazard function.

Another paper related to ours is Ham, Li and Shore-Sheppard (2014): studying the employment dynamics of single mothers in the Survey of Income and Program Participation (SIPP), the authors establish identification of a (discretized) baseline hazard function in a duration model with multiple spells in the presence of seam bias and unobserved heterogeneity.[3] An appealing aspect of their identification strategy is that it can be applied to samples that consist not only of newly sampled ("fresh"), but also of left-censored spells, a common feature of labor market history survey data.[4] The key difference between the identification strategy of Ham, Li and Shore-Sheppard (2014) and ours is that they require at least two measurements of the same duration (collected from different survey waves), where only one is affected by seam bias. Since we do not have multiple measurements at disposal in our data, their methodology cannot be applied in our context.

The identification strategy of our paper is based on a set of minimal assumptions on the shape of the discretized hazard function. That is, neither do we require administrative data nor do we make assumptions about the validity of observations closer to the interview date.[5] Instead, our identification strategy requires, as key ingredients, the existence of at least one correctly reported duration point and the presence of some flat segment(s) in the baseline hazard, which includes this correctly reported duration point. The length of the flat segment(s) required depends on the complexity of the heaping process.

Heuristically speaking, the constant part of the baseline hazard enables us to identify the parameters of the heaping process, i.e. the probability of rounding to a heaped value, 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 have to come from a specific data set. For instance, in our empirical example we partially rely on information from a verbal autopsy report on neonatal mortality in Uttar Pradesh, the most populated Indian state, which suggests that assuming a flat hazard segment towards 18 days is a relatively plausible assumption. The likelihood function is then constructed down-weighting the contribution of heaped durations and over-weighting the contribution of 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.[6]

[3]Seam bias, another common form of measurement error in multi-wave survey data, is characterized by an over-reported frequency of spell failures at the seam of different survey waves.

[4]In particular, the authors show that identification from left-censored spells can be obtained without restricting duration dependence by imposing that misreporting parameters are identical for left-censored and for newly begun spells.

[5]In fact, we are unable to use durations sampled closer to the interview date as a 'validation sample' of 'correctly' reported observations as these durations exhibit similar heaping patterns as the ones that lie further away from the interview date. This suggests that heaping in our data is not mainly driven by recall error.

[6]We rely on a parametric specification for the unobserved heterogeneity distribution to obtain a closed form for the likelihood function. However, 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. See for instance Bierens (2008), Hausman and Woutersen (2014), or Burda, Harding, Hausman (2014)

Moreover, in the supplementary material of the paper we provide an informal discussion on how our methodology can be extended to identify the parameters of different discrete duration models (e.g., Han and Hausman, 1990; Sueyoshi, 1995) as well as of standard ordered choice models. Since ordered choice models are also used in the analysis of count data such as number of doctor visits or cigarette consumption, where heaping is often a feature of the data, we deem this an important extension of our theoretical identification results.

Finally, we derive a number of specification tests that allow applied researchers to check the validity of different heaping mechanisms as well as model specifications. The tests are based on the likelihood ratio statistic and can be straightforwardly computed. However, when some of the heaping parameters lie on the boundary of the parameter space, the limiting distribution of the statistic under the null hypothesis follows a mixture of standard χ^2-distributions, and critical values can be difficult to construct. Thus, we establish the first order validity of critical values based on the 'M out of N' bootstrap.

The rest of this paper is organized as follows. Section 2 describes the setup and the heaping model we consider. As a main result, we provide a set of sufficient conditions for the identification of the parameters of a discrete time duration model. Section 3 derives the likelihood function and establishes the asymptotic properties of the maximum likelihood estimator. Since we do not impose a strictly positive probability of rounding, we account for the possibility of parameters on the boundary of the parameter space (Andrews, 1999). Section 4 outlines how to conduct inference in our context. More specifically, the paper proposes two specification tests to detect whether heaping matters in a statistical sense and, if it matters, to determine whether any of the heaping parameters lie on the boundary. Section 5 then extends the setup of Sections 3 and 4 to analyze the effects of potential policy changes on duration outcomes when inference is hampered by a change in the reporting behavior over time.[7] That is, we discuss the possibility that treatment not only affects the parameters of interest, but also the reporting behavior (and thus the pattern of heaping). We also develop various specification tests to verify these conjectures statistically. Finally, Section 6 contains the empirical example and reports our findings. First, a specification test and model estimates indicate that heaping plays an important role in our data. Second, we do not find evidence for an increase in survival probability of babies born in districts that were treated. However, since our analysis was conducted using only babies born in districts that were eventually treated, it remains to be established whether the actual treatment effect exhibits a similar pattern, too. Section 7 concludes. Figures, graphs, and tables can be found in Appendix I, while Appendix II contains the proofs of our main technical results (Proposition 1 and Theorem 2). All proofs of Sections 4 and 5, some additional specification tests (together with their proofs), instructions on how to implement the bootstrap procedure as well as other supporting material have been collected in the supplementary material.

As a final remark on notation, note that lower case letters are employed throughout the paper to denote random variables as well as their realizations. Also, let $K = \dim(x)$ denote the dimension of any $K \times 1$ vector x with $K \geq 1$, and define $\mathbb{R} = (-\infty, \infty)$ and $\mathbb{R}_+ = (0, \infty)$.

---

for recent advances in dealing with unobserved heterogeneity in duration models with correctly reported observations.

[7] A motivating example is given at the beginning of Section 5.

2 Identification

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

Let τ_i^* be a continuous, non-negative random variable denoting the (continuous) duration time of individual i. The associated hazard function at time τ^* is given by:

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

We parameterize this 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. In most empirical studies, however, time is observed on a discrete scale, e.g., in days in the illustration of Section 6. Thus, in the following, we will assume that a continuous duration $\tau_i^* \in [\tau, \tau+1)$ is recorded as τ, where τ denotes a discrete time period, so that the sample of (discrete) durations is given by τ_i for $i = 1, \ldots, N$. The discrete time hazard for our model can then be written as:

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

where $\gamma(\tau) = \ln \int_\tau^{\tau+1} \lambda_0(s) ds$. Due to misreporting, however, the researcher does not observe τ_i directly, but t_i, a potentially mismeasured version of it.

More specifically, the form of misreporting we address in this paper is referred to as "heaping" in the literature and describes the phenomenon of observing an over- and underreporting of failures at certain time periods. Formally, assume that excessive concentrations of reported failures occur at time period h^* and at multiples $j \cdot h^*$ with $j = 0, 1, \ldots, \overline{j}$ and $\overline{j}$ a finite, non-negative integer denoting the maximum number of heaps in the data.[8] For instance, in the example of the 28 day neonatal mortality period, if reported exits are heaped at values that are multiples of 5 periods and the last heap occurs at period 25, then $h^* = 5$ and $j = 1, 2, 3, 4, 5$. A stylized version of this case has been illustrated in Figure 1 of Appendix I.

Next, to identify the baseline hazard from possibly misreported observations, we impose a structure on the heaping process: let $\overline{r}$ denote the maximum number of time periods that a duration can be rounded to, e.g. with heaps at 5 and 10 days in the stylized example above, $\overline{r}$ is set equal to 2 so that two periods to the right and to the left of each heap point are associated with that heap.

[8] Note that the equal distance between the heap points is a notational simplification, which could straightforwardly be relaxed at the cost of further notation. Likewise, we ignore the possibility of heaping at time period zero, which would again complicate notation unnecessarily by introducing asymmetries into the heaping mechanism (since only time periods to the right can be associated with a heap at time period zero).

Furthermore, assume that all durations τ_i, $i \in 1, \ldots, N$, fall into the set $\mathcal{D} = \{0, 1, \ldots, \overline{\tau}-1\}$, where $\overline{\tau}$ is some finite, positive integer and $(\overline{\tau} - 1)$ represents the maximum observed time period. We assume that all heaping is to observed duration points only.[9] Then, denote the set of
(i) heaping points as:

$$\mathcal{D}^{\mathcal{H}} = \left\{\tau : \tau = jh^*, j = 1, 2, ..., \overline{j},\ \overline{j}h^* < \overline{\tau}\right\};$$

(ii) points that may be rounded up as:

$$\mathcal{D}^{\mathcal{H}_{-l}} = \left\{\tau : \tau = jh^* - l, j = 1, 2, ..., \overline{j},\ \overline{j}h^* - l < \overline{\tau}\right\};$$

and (iii) points that may be rounded down as:

$$\mathcal{D}^{\mathcal{H}_{+l}} = \left\{\tau : \tau = jh^* + l, j = 1, 2, ..., \overline{j},\ \overline{j}h^* + l < \overline{\tau}\right\};$$

for $l = 1, ..., \overline{r}$. See Figure 2 in Appendix I for an illustration of the case where $h^* = 5$, $\overline{j} = 2$, and $\overline{r} = 2$.

Moreover, assume that whenever the true duration τ_i falls onto 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..[10] 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 $\tau_i \in \mathcal{D}^{\mathcal{H}_{-l}}$. 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 the example from above, a possible mechanism is for instance that a reported duration of say 10 days includes true durations of 11 and 12 (8 and 9), which have been rounded up (down) to 10 days (see again Figure 2 in Appendix I). The ps and the qs give the probabilities of these roundings.

Next, we outline our key identification assumption:

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

Assumptions H(i)-(iv) are crucial for the identification of the baseline parameters and our identification result in Proposition 1 below. More specifically, H(i)-(ii) imply that duration points are only associated with one (the nearest) heaping point. In our example, this rules out that true durations of 7 days are, for instance, rounded to 5 *and* 10 days simultaneously. We note, however, that our setup could also accommodate more complex heaping mechanisms provided that the constant segment of the hazard required by H(iii) is sufficiently

[9]For reasons of clarity, we will introduce censoring only in the next section.

[10]Note that the assumption that true durations τ_i, which fall onto a heap point, are correctly observed in principle rules out scenarios where 'standard' measurement error (e.g., due to recall error) affects *all* durations. We acknowledge that this might be too strong of an assumption for some empirical settings, and in fact think of our model framework as an admissible approximation only in applications where this general form of measurement error is negligible relative to the effects of heaping. However, as pointed out in Footnote 4, we believe this assumption to be relatively plausible in the context of our empirical illustration.

large (see below). Moreover, in Section S2 of the supplementary material, we discuss a specification test that allows to discriminate between different (identified) heaping mechanisms. Assumption H(ii) also imposes that durations belonging to either $\mathcal{D}^{\mathcal{H}_{-l}}$ or $\mathcal{D}^{\mathcal{H}_{+l}}$ with $l = 1, \dots, \overline{r}$ have a strictly positive probability to be truthfully reported and to be observed. This assumption, which is essential to identify the parameters $\gamma(k)$ for $1 < k < \overline{\tau}$, is deemed to be a rather mild regularity condition satisfied by most empirical settings. H(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 $j \cdot h^*$, which could for instance apply when heaping is asymmetric. Moreover, in H(iv), it is assumed that $\overline{k}$ is observed without error. That is, $\overline{k} \in \mathcal{D}^{\mathcal{T}}$, where $\mathcal{D}^{\mathcal{T}}$ is the set of truthfully reported durations and defined as the following complement set:

$$\mathcal{D}^{\mathcal{T}} = \left(\left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}} \right) \cup \left(\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}} \right) \cup \mathcal{D}^{\mathcal{H}} \right)^c .$$

Note that the set $\mathcal{D}^{\mathcal{T}}$ may not only contain $\overline{k}$, but also other duration points $\tau < \overline{\tau}$, which are known to be correctly observed and do not lie in the union of $\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{-l}}$, $\cup_{l=1}^{\overline{r}} \mathcal{D}^{\mathcal{H}_{+l}}$, and $\mathcal{D}^{\mathcal{H}}$. Finally, we emphasize again that Assumptions H(iii)-(iv) are stronger than required as it would in principle suffice for the hazard to be constant over separate regions, not necessarily adjacent to each other, as long as these regions cover $\mathcal{D}^{\mathcal{H}}$, $\mathcal{D}^{\mathcal{H}_{-l}}$, and $\mathcal{D}^{\mathcal{H}_{+l}}$, $l = 1, \dots, \overline{r}$.[11]

Heuristically, the assumption that the hazard is constant over a set of time periods, which includes (at least) a known true value, enables us to uniquely identify the γ parameter of the correctly reported time period as well as the parameters modeling the heaping process, i.e. the p's and the q's, in this region. Subsequently, we can use these identified probability parameters to pin down the rest of the baseline hazard parameters.

Before stating our main identification result, we need to define some more notation, which will be used in the following. Let $\underline{\theta} = \{\beta', \gamma'\}'$ with $\gamma = \{\gamma(0), \gamma(1), \dots, \gamma(\overline{\tau} - 1)\}'$ and define the probability of survival at least until time period $\tau < \overline{\tau}$ in the absence of misreporting as:

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

[11] Moreover, note that there are different alterations of Assumption H that yield identification of the parameters of interest. For instance, dropping assumption H(iv), one could still obtain the main identification result of Proposition 1 if $\gamma(k) = \overline{\gamma}$ for all $j \cdot h^* - \overline{r} \leq k \leq j \cdot h^* + \overline{r}$ in H(iii). In other words, the $\gamma(\cdot)$ parameters to the left *as well as* to the right of the heap $j \cdot h^*$ are required to be constant and identical if H(iv) is dropped.

where $v_i \equiv \exp(u_i)$. The probability for an exit event in $\tau_i < \overline{\tau}$ is:

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

Here, $f_i\left(\tau|z_i,u_i,\underline{\theta}\right)$ 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 function (see next section). Hereafter, let

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

with t_i denoting the discrete *reported* duration. It is immediate to see the following for the four cases:

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

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

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

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

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

In summary, there are four different probabilities of exit events depending on whether the reported duration t_i is in $\mathcal{D}^{\mathcal{T}}$, $\mathcal{D}^{\mathcal{H}_{-l}}$, $\mathcal{D}^{\mathcal{H}_{+l}}$, or $\mathcal{D}^{\mathcal{H}}$ respectively. Next, in order to obtain the unconditional versions of these probabilities, we introduce the following assumption on the unobserved heterogeneity distribution:

Assumption U:
(i) v_i is identically and independently distributed;
(ii) The density of v is gamma with unit mean and variance σ^{-1};
(iii) v_i is independent of z_i.

Assumptions U(i)-(iii) allow to integrate out unobserved heterogeneity and so to obtain the unconditional versions of the above probabilities. The parametric choice in Assumption U(ii) on the other hand ensures that these unconditional probabilities have a closed form expression, 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 in principle only require some regularity condition of the form $\mathbb{E}[v_i] < \infty$ (e.g., see Ridder and Woutersen, 2003). That is, while the gamma density choice might appear overly restrictive at first sight, we note that U(ii) can often be rationalized 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, we impose the following assumption on the observed heterogeneity distribution:

Assumption Z: The support of at least one element z_{1i} of z_i, say $\mathcal{S}_{z1}$, contains at least two values and the corresponding element of β is non-zero. Moreover, the full support of z_i, $\mathcal{S}_z$, contains the zero vector.

Assumption Z is standard in the literature on identification of MPH models (cf. Elbers and Ridder, 1982; Ridder and Woutersen, 2003) and requires a minimum amount of variation in the covariates z_i to identify β.

Using assumption U, the unconditional probabilities in case (I) above are given by:

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

where the last equality uses the fact that there is a closed form expression under U(ii) (e.g., see Meyer (1990, p. 770)). Moreover, since the integral is a linear operator the probabilities for the cases (II) to (IV) can be derived accordingly. Then, we obtain the following identification result:

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

The proof is provided in Appendix II and its heuristics are explained in the first of the following two remarks:

Remark 2.1: The proof of Proposition 1 is based on establishing a one to one relationship between exit probabilities and the baseline hazard (and the other model) parameters. Given Assumptions U, H(iii)-(iv) and an argument from Heckman and Singer (1984), we 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(iii), we identify the heaping probabilities ps and qs. Finally, using H(i), we sequentially identify all remaining γ parameters.[12]

Remark 2.2: In Section S3 of the supplementary material we show that the identification idea of Proposition 1 can also be applied or modified to apply to other discrete duration (e.g., Han and Hausman, 1990; Sueyoshi, 1995) and to standard ordered choice models, thus amplifying the applicability of Proposition 1 and the remaining results of the paper.

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

3 Estimation

Our next goal is to obtain consistent estimators for $\theta = \{\underline{\theta}', \sigma, p_1, \ldots, p_{\overline{r}}, q_1, \ldots, q_{\overline{r}}\}'$ from the possibly misreported durations. Before setting up the likelihood function drawing from the derivations of the previous section for truthfully and misreported durations, we introduce censoring into our setup:

Assumption C:
(i) Durations are censored at a fixed time $\overline{\tau} > \overline{j}h^* + \overline{r}$ 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 C(i), which has been made in view of the illustration in Section 6, could be straightforwardly generalized to allow for varying censoring times across individuals (random censoring) as long as C(ii) remains satisfied.[13] Furthermore, under Assumption C, the identification result established in Proposition 1 carries through to the censored case by defining a set $\mathcal{D}^{\mathcal{C}}$ for observations censored at $\overline{\tau}$, which does not overlap with $\mathcal{D}^{\mathcal{H}+l}$, $l = \overline{r}$.

Next, let δ_i be an indicator equal to one if the observation does not lie in $\mathcal{D}^{\mathcal{C}}$ and zero otherwise. Then, given Assumptions U, Z, and C and the definition of $\phi_i(\cdot)$ from cases (I) to (IV) in Section 2, let the likelihood function be:

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

and so

$$l_N(\theta) = \ln L_N(\theta) = \sum_{i=1}^{N} \ln \int \Big\{ \phi_i(t|z_i, v)^{\delta_i} S_i(t|z_i, v)^{(1-\delta_i)} \Big\} g(v; \sigma) \mathrm{d}v.$$

Given the definition of $\phi_i(t|z_i, v)$ and cases (I) through (IV) in Section 2, it is clear that the (log) likelihood function down-weights the contribution of heaped durations and over-weights the contribution of non heaped durations. Thus

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

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

where $\Theta \subset \mathbb{R}^{K_\beta + K_\gamma} \times \mathbb{R}_+ \times [0, 1)^{2\overline{r}}$ denotes the parameter space with $K_\beta = \dim(\beta)$ and $K_\gamma = \dim(\gamma)$.

Assumption D:
(i) $(\tau_i, x_i')'$, $i = 1, \ldots, N$, are i.i.d. random variables that take values in a subset of the product space $\mathbb{R}_+ \times \mathbb{R}^{K_x}$.
(ii) $\mathbb{E}[\tau_i^4] < \infty$.
(iii) For all $d = 1, ..., \overline{\tau} - 1$, $\frac{1}{N} \sum_i 1\{\tau_i = d\} \xrightarrow{p} \Pr[\tau_i = d] > 0$.

[13] That is, a similar structure to cases (I) to (IV) in Section 2 could be used to determine the survival probabilities of censored spells (and thus their contribution to the likelihood). Also note that $\overline{\tau}$ does not necessarily have to be greater than $\overline{\tau} > \overline{j}h^* + \overline{r}$ as long as it does not belong to $\mathcal{D}^{\mathcal{H}}$ or $\mathcal{D}^{\mathcal{H}+l}$ with $l = 1, \ldots, \overline{r}$.

Together with Assumption H(ii), Assumption D(iii) ensures that asymptotically we observe exits in each time period until $\overline{\tau}-1$, which in turn allows to estimate the associated baseline hazard parameters consistently.[14] We now establish the asymptotic properties of $\widehat{\theta}_N$.

Theorem 2: *Let Assumptions* H,Z,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 $\mathcal{I}^\dagger = E\left(\left(-\nabla^2_{\theta\theta} l_N(\theta)/N\right)|_{\theta=\theta^\dagger}\right)$, $G \sim N\left(0, \mathcal{I}^{\dagger-1}\right)$, *and* Ψ *being a cone in* $\mathbb{R}^{K_\beta+K_\gamma} \times \mathbb{R}_+ \times [0,1)^{2\overline{\tau}}$.[15]

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

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

Theorem 2 establishes consistency of $\hat{\theta}_N$ in part (i) and its limiting distribution in parts (ii) and (iii). Its proof can be found in Appendix II. Note that the limiting distribution of the estimator depends on whether some heaping probability parameters lie on the boundary of the parameter space or not. That is, if one or more of the "true" probability parameters are equal to zero, the limiting distribution of $\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right)$ is no longer normal as the information matrix $\mathcal{I}^{\dagger-1}$ is not block diagonal in general, but takes the form in part (ii) (cf. Andrews, 1999).

4 Inference

Inference hinges on the limiting distribution of $\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right)$. However, as outlined above, if some of the heaping probability parameters lie on the boundary of the parameter space and the asymptotic distribution of $\sqrt{N}\left(\widehat{\theta}_N - \theta^\dagger\right)$ is no longer normal, inference on the baseline hazard and other model parameters becomes more complicated. A solution in this case is the use of subsampling methods or, more specifically, the M out of N bootstrap (Andrews, 1999, 2000).[16]

In the following, we propose two specification tests (i) to detect whether heaping matters in a statistical sense ($\mathbf{H}^{\pi_1}$), and, if it matters, (ii) to discriminate between the general case in Theorem 2(ii) that allows for probability parameters on the boundary and the special case in Theorem 2(iii) without parameters on the boundary ($\mathbf{H}^{\pi_2}$). That is, while the first test

[14] Even though, due to the model complexity, an evaluation of the finite sample performance of our methodology is beyond the scope of this paper, it is certainly the case that sufficiently large samples are required to observe enough exits in each time period. Moreover, in particular towards the tail of the baseline function, coarse data observations may prevent the identification of the heaping parameters in practice.

[15] Ψ is a cone in $\mathbb{R}^K$, if for $a > 0$, $\psi \in \Psi$ implies $a\psi \in \Psi$.

[16] Section S3 in the supplementary material outlines the implementation of the M out of N bootstrap in our setup.

helps to determine whether the specified heaping model is indeed preferred over a standard duration model that does not account for heaping, the second test allows to decide whether inference in fact ought to be based on subsampling methods.

Thus, collecting all heaping parameters in the vector π with $\pi = \{p_1, \dots, p_{\overline{r}}, q_1, \dots, q_{\overline{r}}\}'$ and $\theta = \{\underline{\theta}', \sigma, \pi'\}'$, the first test examines the existence of heaping effects through:

$\mathbf{H}^{\pi_1}$:

$$H_0^{\pi_1} : p_1 = ... = p_{\overline{r}} = q_1 = ... = q_{\overline{r}} = 0$$

vs

$$H_A^{\pi_1} : p_l > 0 \text{ and/or } q_l > 0$$

for some $l = 1, ..., \overline{r}$. Defining $\Theta_0 \subset \mathbb{R}^{K_\beta + K_\gamma} \times \mathbb{R}_+ \times \mathbf{0}^{2\overline{r}}$ to be the parameters space under $H_0^{\pi_1}$, where $\mathbf{0}^{2\overline{r}}$ denotes a zero vector of dimension $2\overline{r}$, let:

$$l_N\left(\widetilde{\theta}_N\right) = \sup_{\theta \in \Theta_0} l_N(\theta) \quad \text{and} \quad l_N\left(\widehat{\theta}_N\right) = \sup_{\theta \in \Theta} l_N(\theta).$$

Then, a Likelihood Ratio (LR) test can be set up as:

$$LR_N = -2\left(l_N\left(\widetilde{\theta}_N\right) - l_N\left(\widehat{\theta}_N\right)\right).$$

The following proposition derives the asymptotic distribution of LR_N under the null hypothesis and establishes consistency against fixed general alternatives.

Proposition 3: *Let Assumptions* H,U,C, Z, *and* D *hold. Then, for* H^{π_1}, *we have:*
(i) Under $H_0^{\pi_1}$,

$$LR_N \xrightarrow{d} \min_{\psi \in \Psi_0} (\psi - G)' \mathcal{I}_{\theta_0} (\psi - G) - \min_{\psi \in \Psi} (\psi - G)' \mathcal{I}_{\theta_0} (\psi - G),$$

where $G \sim N\left(0, \mathcal{I}_{\theta_0}^{-1}\right)$, Ψ *is a cone in* $\mathbb{R}^{K_\beta + K_\gamma} \times \mathbb{R}_+ \times [0,1)^{2\overline{r}}$, and Ψ_0 *is a cone in* $\mathbb{R}^{K_\beta + K_\gamma} \times \mathbb{R}_+ \times \mathbf{0}^{2\overline{r}}$.
(ii) Under $H_A^{\pi_1}$, *there exists* $\varepsilon > 0$ *such that*

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

The proof of Proposition 3 (as well as the proofs of all other Propositions of this and the next section) can be found in Section S1 of the supplementary material. The following comments are worth noting:

Remark 4.1: The statement in part (i) does not necessarily require that the heaping structure, or the model in general, is correctly specified. If the model is not correctly specified, the information matrix equality does not hold in general. In this case, $\min_{\psi \in \Psi_0} (\psi - G)' \mathcal{I}_{\theta_0} (\psi - G) - \min_{\psi \in \Psi} (\psi - G)' \mathcal{I}_{\theta_0} (\psi - G)$ no longer follows a $\overline{\chi}^2$-distribution with $2\overline{r}$ degrees of freedom, the limiting distribution of LR_N (e.g., see Chapter 3 in Silvapulle and Sen (2005)). However, since we base inference on M out of N bootstrap critical values in the illustration of Section 6, this issue does not arise in the context of our paper (see Remark 4.2).

Remark 4.2: As noted in the previous remark, the limiting distribution of LR_N is a $\overline{\chi}^2$-distribution, a mixture of standard χ^2-distributions. The weights of this mixture can be computed by simulation, and thus, given an estimator for $\mathcal{I}_{N,\theta_0}$, critical values can be constructed. However, when Θ is of high dimension, both the estimation of $\mathcal{I}_{N,\theta_0}$ as well as of the weights is rather cumbersome. For this reason, a computationally more convenient strategy is to use critical values based on the M out of N bootstrap. That is, let $l_i(\theta)$ be the contribution of the i-th duration to the likelihood $l_N(\theta)$. Let I_j, $j = 1, ..., M$ be M independent draws from a discrete uniform distribution on $[1, N]$. We then make M draws with replacement from $(l_1(\theta), ..., l_M(\theta))$ to obtain $(l_{I_1}(\theta), ..., l_{I_M}(\theta)) = (l_1^*(\theta), ..., l_M^*(\theta))$, $M = o(N)$. Then, let

$$\widehat{\theta}_M^* = \arg\max_{\theta\in\Theta} l_M^*(\theta) \text{ and } \widetilde{\theta}_M^* = \arg\max_{\theta\in\Theta_0} l_M^*(\theta)$$

and

$$LR_M^* = -2\left(l_M^*\left(\widetilde{\theta}_M^*\right) - l_M^*\left(\widehat{\theta}_M^*\right)\right).$$

Also, let $c^*_{(1-\alpha),M,N,B}$ denote the $(1-\alpha)-$th percentile of the empirical distribution of $LR_M^{*(j)}$, $j = 1, ..., B$, where the superscript (j) denotes the $j-$th bootstrap replication. We obtain the following result in analogy to Proposition 3:

Proposition 3*: *Let Assumptions* H,U,Z,C *and* D *hold, and let* $M/N \to 0$. *Then, for* H^{π_1}, *we have:*

(i) Under $H_0^{\pi_1}$,

$$\lim_{M,N,B\to\infty} P\left(LR_N > c^*_{(1-\alpha),M,N,B}\right) = \alpha$$

(ii) Under $H_A^{\pi_1}$,

$$\lim_{M,N,B\to\infty} P\left(LR_N > c^*_{(1-\alpha),M,N,B}\right) = 1.$$

The proof of Proposition 3* can be found in Section S1 of the supplementary material.

Remark 4.3: In Section S2.1 of the supplementary material, we discuss an additional specification test that allows to discriminate between different heaping mechanisms (in our model). This test, based on the Kullback Leibler Information Criterion (KLIC; Vuong, 1989), is useful to compare the validity of different potential rounding mechanisms. For instance, suppose that Assumption H holds with $h^* = 5$, $\overline{j} = 5$, $\overline{r} = 2$. In this case, we know that, by Proposition 1, p_1, p_2, q_1, q_2 are uniquely point identified. Then, a potential interest could be to choose between two different rounding schemes, e.g. Model 1 with $p_2 = q_2 = 0$ and $p_1 \neq q_1$, or Model 2 with p_2 and q_2 different from zero, but $p_1 = q_1$ and $p_2 = q_2$. The test in the supplementary material allows for such a model comparison.

Remark 4.4: In Section S2.2 of the supplementary material, we show that a test of the null hypothesis that the elements of β and/ or σ, and/ or γ are the same in a model with and without heaping parameters is equivalent to the test in Proposition 3.

The second specification test examines whether all heaping parameters lie inside the parameter space, which in turn allows inference based on asymptotic normality. That is, the null hypothesis of the test is that at least one rounding parameter is equal to zero versus the alternative that none is zero (and thus no boundary problem exists). Formally, 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 hypothesis:

$\mathbf{H}^{\pi_2}$:

$$H_0^{\pi_2} = \left(\cup_{j=1}^{\overline{r}} H_{p,0}^{(j)}\right) \cup \left(\cup_{j=1}^{\overline{r}} H_{q,0}^{(j)}\right)$$

vs

$$H_A^{\pi_2} = \left(\cap_{j=1}^{\overline{r}} H_{p,A}^{(j)}\right) \cap \left(\cap_{j=1}^{\overline{r}} H_{q,A}^{(j)}\right),$$

so that under $H_A^{\pi_2}$ all p's and q's are strictly positive. To discriminate between $H_0^{\pi_2}$ and $H_A^{\pi_2}$, we apply the Intersection-Union principle (IUP), see e.g. Chapter 5 in Silvapulle and Sen (2005). According to the IUP, we only reject $H_0^{\pi_2}$ at level α if all single null hypotheses $H_{p,0}^{(j)}$ and $H_{q,0}^{(j)}$ are rejected at level α.

Let

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

where $\widehat{\mathcal{I}}_N^{1/2}\widehat{\mathcal{I}}_N^{1/2} = \widehat{\mathcal{I}}_N$, $\widehat{\mathcal{I}}_N = \frac{1}{N}\nabla_\theta l_N(\widehat{\theta})\nabla_\theta' l_N(\widehat{\theta})$ and $\widehat{\mathcal{I}}_{p_j p_j,N}, \widehat{\mathcal{I}}_{q_j q_j,N}$ are the corresponding entries. Also, let

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

where Z denotes a standard normal random variable. We now introduce a rule to discriminate between $H_0^{\pi_2}$ and $H_A^{\pi_2}$.

Rule IUP-PQ: Reject $H_0^{\pi_2}$, if $\max_{j=1,\ldots,\overline{r}} \{PV_{p,j}, PV_{q,j}\} < \alpha$ and do not reject otherwise.

Proposition 4 below establishes that a test based on Rule IUP-PQ has correct asymptotic size and power against fixed general alternatives.

Proposition 4: *Let Assumptions* H,U,Z,C, *and* D *hold. Then, Rule IUP-PQ ensures that for* H^{π_2}

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\pi_2} | H_0^{\pi_2} \text{ true}\right) \leq \alpha$$

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\pi_2} | H_0^{\pi_2} \text{ false}\right) = 1.$$

Thus, if one rejects $H_0^{\pi_2}$, inference can be based on asymptotic normality, while failure to reject $H_0^{\pi_2}$ requires the use of subsampling methods as outlined before.

5 Policy Analysis Under Heaping

In many empirical studies, the focus of interest lies on the analysis of the effects of certain policy changes on duration outcomes. In some instances, however, this analysis may be hampered by a change in the reporting behavior over time due to the implementation of those same policies, which could confound any observed effect. For instance, in the illustration of Section 6, any potential effect of the JSY program on neonatal mortality could be masked by a change in the (average) accuracy with which mothers report birth dates of their children. That is, if a direct consequence of the program was that more women delivered in health facilities, which issued birth certificates when discharging the mothers, records could, on average, be more accurate than before (due to incorporation of this information either by the interviewers or by the mothers themselves).

To model the effect of such a policy change on the hazard as well as on the heaping probability parameters, consider the following amendments to the setup in Sections 2 and 3: let D_i denote a dummy variable equal to one if the duration is sampled after the introduction of the policy, and zero otherwise.[17] Let the probability of an exit event at τ in the absence of misreporting be denoted as:

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

where $\underline{\vartheta}=\{\theta',\gamma_2'\}'$ with θ defined at the beginning of Section 3 and $\gamma_2=\{\gamma^{(2)}(0),\ldots,\gamma^{(2)}(\overline{\tau}-1)\}'$. The coefficient of D_i, $\gamma^{(2)}(\cdot)$, is defined analogously to $\gamma(\cdot)$ and thus measures the change w.r.t. $\gamma(\cdot)$ after the policy introduction. The contribution of a non-truthfully reported duration is defined in analogy to $\phi_i\left(t|z_i,u_i,\underline{\theta}\right)$ in Section 2, say $\widetilde{\phi}_i\left(t|z_i,u_i,\underline{\vartheta}\right)$. Thus, (I) for any $t_i=t\in\mathcal{D}^{\mathcal{H}_{-l}}$,

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

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

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

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

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

where the definition of $p_l^{(2)}$ and $q_l^{(2)}$, $l=1,\ldots,\overline{r}$, is immediate. The above specification allows for a potential "structural break" in the heaping parameters with the policy introduction. Thus, in the sequel we shall need

[17] We assume that D_i, similar to the observed characteristics z_i, is measured without error. This assumption, albeit problematic in some settings, does not appear to be implausible in the context of our illustration as birth information, which is used to determine D_i, stems from a different variable than the information about the reported duration t_i (see next section).

Assumption D':
(i) $(\tau_i, x_i')'$, $i = 1, \ldots, N$, are independent but not identically distributed. random variables that take values in a subset of the product space $\mathbb{R}_+ \times \mathbb{R}^{K_x}$.
(ii) $\mathbb{E}[\tau_i^{4(1+\delta)}] < \infty$ for $\delta > 0$.
(iii) As in Assumption D.

Note that durations are not necessarily identically distributed, as their distribution may differ depending on whether they occur before or after the introduction of the policy. Finally, with slight abuse of notation, let $\vartheta = \{\underline{\vartheta}', \pi_2'\}'$ with $\pi_2 = \{p_1^{(2)}, \ldots, p_{\overline{r}}^{(2)}, q_1^{(2)}, \ldots, q_{\overline{r}}^{(2)}\}'$ and $\pi_1 = \{p_1, \ldots, p_{\overline{r}}, q_1, \ldots, q_{\overline{r}}\}' \in \Pi_1 \subseteq [0,1)^{2\overline{r}}$ in the following. Also note that

$$\pi_2 \in \Pi_2 = [-\Pi_1, \mathbf{1} - \Pi_1),$$

where $\mathbf{1}$ denotes a $2\overline{r}$ dimensional vector of ones. $\widetilde{l}_N(\vartheta)$ is defined as $l_N(\theta)$ replacing $\phi_i\left(t|z_i, u_i, \underline{\theta}\right)$ by $\widetilde{\phi}_i\left(t|z_i, u_i, \underline{\vartheta}\right)$. Then, letting

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

and

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

with $\widetilde{\Theta} \subset \Theta \times \mathbb{R}^{K_{\gamma 2}} \times \Pi_2$, where $K_{\gamma 2} = \dim(\gamma_2)$ and Θ is defined in Section 3, it is straightforward to show that the identification and asymptotic results of Sections 2 and 3 continue to hold.

We discuss two different types of specification tests in the following: first, to understand if a policy change did affect the reporting behavior and thus also the heaping mechanism, we develop a test for the null hypothesis of no change in the heaping probability parameters after the policy introduction vs. the alternative of a change in at least some rounding parameters ($\mathbf{H}^{\pi_3}$).

Second, we propose a set of tests to examine whether the policy had indeed any effect on the actual parameters of interest (i.e. the baseline hazard parameters).

In a first step, we thus test the null hypothesis that the policy did not lead to any shift in the baseline hazard parameters versus the alternative that it did in fact lead to a uniform (over the observation period) downward shift of those parameters ($\mathbf{H}^{\gamma_1}$). If we fail to reject this null hypothesis, we proceed, in a second step, by testing whether no baseline hazard parameter changed over the observation period versus a change of at least some parameter ($\mathbf{H}^{\gamma_2}$). Together, these latter tests provide an overall picture of the effectiveness of the policy and whether it had the desired effect on the outcome durations (e.g., whether the JSY program effectively reduced neonatal mortality in districts where it was implemented).

We start with the test for a change in the heaping parameters and test:

$\mathbf{H}^{\pi_3}$:

$$H_0^{\pi_3} : p_1^{(2)} = ... = p_{\overline{r}}^{(2)} = q_1^{(2)} = ... = q_{\overline{r}}^{(2)} = 0$$

vs

$$H_A^{\pi_3} : p_l^{(2)} \neq 0 \text{ and/or } q_l^{(2)} \neq 0$$

for some $l = 1, ..., \overline{r}$. Note that $H_A^{\pi_3}$ is stated as a two-sided alternative. However, given that $\Pi_2 = [-\Pi_1, \mathbf{1} - \Pi_1)$, if some elements of Π_1 are zero, the corresponding elements of Π_2 under the alternative can only be positive. Next, let $\vartheta^{\pi_3} = \{\underline{\vartheta}', \mathbf{0}^{2\overline{r}\prime}\}'$ and

$$\widehat{\vartheta}_N^{\pi_3} = \arg \max_{\vartheta^{\pi_3} \in \widetilde{\Theta}^{\pi_3}} l_N \left(\vartheta^{\pi_3}\right) \text{ and } \widehat{\vartheta}_N = \arg \max_{\vartheta \in \widetilde{\Theta}} l_N \left(\vartheta\right)$$

as well as

$$\widehat{\vartheta}_M^{*\pi_3} = \arg \max_{\vartheta^{\pi_3} \in \widetilde{\Theta}^{\pi_3}} l_M^* \left(\vartheta^{\pi_3}\right) \text{ and } \widehat{\vartheta}_M^* = \arg \max_{\vartheta \in \widetilde{\Theta}} l_M \left(\vartheta\right),$$

where $\widetilde{\Theta}^{\pi_3} \subset \Theta \times \mathbb{R}^{K_{\gamma_2}} \times \mathbf{0}^{2\overline{r}}$ and $l_M^*(.)$ is defined as in Section 3 above. Then, replacing Assumption D by Assumption D$'$, and letting

$$LR_N^{\pi 3} = -2\left(l_N\left(\widehat{\vartheta}_N^{\pi 3}\right) - l_N\left(\widehat{\vartheta}_N\right)\right)$$

and

$$LR_M^{*\pi 3} = -2\left(l_M^*\left(\widehat{\vartheta}_M^{*\pi 3}\right) - l_M\left(\widehat{\vartheta}_M^*\right)\right),$$

the statements of Proposition 3 and Proposition 3* in Section 4 continue to hold.

Remark 5.1: Note that, if all components of Π_1 are strictly positive, then there is no parameter on the boundary and, under $H_0^{\pi_3}$, $LR_N^{\pi 3} \xrightarrow{d} \chi^2_{2\overline{r}}$. In this case, we do not need to rely on M of out N boostrap critical values.

Next, we turn to the specification tests aimed at detecting (potential) changes in the baseline hazard parameters due to a policy change. Our first test examines whether the baseline hazard parameters declined uniformly over the observation period. That is:

$$H_0^{\gamma_1} : \max\left\{\gamma^{(2)}(0), \gamma^{(2)}(1), ..., \gamma^{(2)}(\overline{\tau} - 1)\right\} \geq 0$$

vs

$$H_A^{\gamma_1} : \max\left\{\gamma^{(2)}(0), \gamma^{(2)}(1), ..., \gamma^{(2)}(\overline{\tau} - 1)\right\} < 0.$$

The null hypothesis states that the baseline hazard function has either increased or not changed over at least one duration point. On the other hand, under the alternative, the policy has reduced neonatal mortality over all periods considered, i.e. over every day the baseline hazard has decreased. Note that $H_A^{\gamma_1}$ implies

$$\widetilde{H}_A^{\gamma_1} : \max_{J \leq \overline{\tau} - 1} \left\{\sum_{j=0}^{J} \gamma^{(2)}(j)\right\} < 0$$

while $\widetilde{H}_A^{\gamma_1}$ does not necessarily imply $H_A^{\gamma_1}$. Thus, rejection of $H_0^{\gamma_1}$ is a sufficient, but not a necessary condition for a uniform downward shift of the baseline hazard function. In other words, rejection of the null hypothesis suggests that the policy has lowered the exit risk over the observation period. With slight abuse of notation, we re-state $H_0^{\gamma_1}$ and $H_A^{\gamma_1}$ as:

$\mathbf{H}^{\gamma_1}$:

$$H_0^{\gamma_1} = \cup_{j=1}^{\overline{\tau}-1} H_{\gamma,0}^{(j)}$$

vs

$$H_A^{\gamma_1} = \cap_{j=1}^{\overline{\tau}-1} H_{\gamma,A}^{(j)},$$

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. As in Section 3, we apply again the Intersection Union Principle, IUP. Let:

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

with Z being a standard normal random variable.

Rule IUP-GAMMA2: Reject $H_0^{\gamma_1}$, if $\max_{j=1,\dots,\overline{\tau}-1}\left\{PV_{\gamma_j^{(2)},N}\right\} < \alpha$ and do not reject otherwise.

Proposition 5: *Let Assumptions* H,U,Z,C *and* D' *hold. Then, Rule IUP-GAMMA2 ensures that for* H^{γ_1}

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\gamma_1} | H_0^{\gamma_1} \text{ true}\right) \leq \alpha$$

$$\lim_{N\to\infty} \Pr\left(\text{Reject } H_0^{\gamma_1} | H_0^{\gamma_1} \text{ false}\right) = 1.$$

As outlined previously, rejecting $H_0^{\gamma_1}$ provides evidence in favor of the efficacy of the policy change.[18] If instead we fail to reject $H_0^{\gamma_1}$, a natural step to proceed is to test the null hypothesis that treatment has not changed the exit probability (e.g., the probability of a baby dying) in any of the first $(\overline{\tau}-1)$ periods against the alternative that over at least one period the exit probability decreased. Formally:

$\mathbf{H}^{\gamma_2}$:

$$H_0^{\gamma_2} : \gamma^{(2)}(0) = \gamma^{(2)}(1) ... = \gamma^{(2)}\left(\overline{\tau}-1\right) = 0$$

vs

$$H_A^{\gamma_2} : \gamma^2(j) < 0 \text{ for some } j = 1, ..., \overline{\tau}-1.$$

Define

$$\vartheta^{\gamma_2} = \{\theta', \mathbf{0}^{K_{\gamma_2}\prime}, \pi_2'\}'$$

and let $\widetilde{\Theta}^{\gamma_2} \subseteq \Theta \times \mathbf{0}^{K_{\gamma_2}} \times \Pi_2$. Then, let

$$\widehat{\vartheta}_N^{\gamma_2} = \arg\max_{\vartheta^{\gamma_2}\in\widetilde{\Theta}^{\gamma_2}} l_N\left(\vartheta^{\gamma_2}\right) \text{ and } \widehat{\vartheta}_N = \arg\max_{\vartheta\in\widetilde{\Theta}} l_N\left(\vartheta\right)$$

and

$$\widehat{\vartheta}_M^{*\gamma_2} = \arg\max_{\vartheta^{\gamma_2}\in\widetilde{\Theta}^{\gamma_2}} l_M^*\left(\vartheta^{\gamma_2}\right) \text{ and } \widehat{\vartheta}_M^* = \arg\max_{\vartheta\in\widetilde{\Theta}} l_M\left(\vartheta\right),$$

where $l_M^*(.)$ defined as in Section 4. Now let

$$LR_N^{\gamma_2} = -2\left(l_N\left(\widehat{\vartheta}_N^{\gamma_2}\right) - l_N\left(\widehat{\vartheta}_N\right)\right)$$

[18] In the example of Section 6, rejecting the null hypothesis would for instance suggest that the JSY program reduced neonatal mortality uniformly over the first month after birth and thus has indeed been a very effective program.

as well as

$$LR_M^{*\gamma_2} = -2\left(l_M^*\left(\widehat{\vartheta}_M^{*\gamma_2}\right) - l_M\left(\widehat{\vartheta}_M^*\right)\right),$$

and replace Assumption D again by Assumption D′. Then, the same statements as in Proposition 3 and Proposition 3* from Section 4 hold.

6 Empirical Illustration

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.[19] DLHS3 (DLHS2) survey collected information from 720,320 (620,107) households residing in 612 (593) districts in 28 (29) states and 6 union-territories (UTs) of India during the period 2007-08 (2002-04). The surveys focused mainly on women and were designed to provide information on maternal and child health along with family planning and other reproductive health services (see Section S4 in the supplementary material for a more detailed description of the survey design and the sample).

The year and month of birth were recorded for all live births. For those children who had died by the time of the interview, the age of the child at the time of death was also collected (in days if the child had died within the first month, in months thereafter). Note however that this age-at-death information is self-reported and thus subject to (potential) reporting error. Finally, the survey provides information on the month and year of interview. We therefore exclude those children who were born within two months of the interview to ensure that all children have had at least one month of exposure.

Janani Suraksha Yojana (JSY) Program: The National Rural Health Mission (NRHM) launched the program Janani Suraksha Yojana (JSY) in April 2005 replacing an earlier program (National Maternity Benefit Scheme (NMBS)) aimed at the provision of better diet. The objective of the JSY was to reduce maternal and neonatal mortality by promoting institutional delivery. 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 initial cash assistance to eligible women for delivery care 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.

Sample and Variables: We do not have information on when and which districts implemented the program. We follow Lim et al. (2010) and Powell-Jackson et al. (2015) 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

[19] 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`.

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[20] 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.[21]

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 address this issue by including a district level variable measuring neonatal mortality in 2000 and by conducting 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. We therefore compare durations from districts that were recorded as treated at the start date of the corresponding duration with those from districts that had not yet been treated by the time the duration had begun (control group).

The object of interest is the deaths within the first 28 days after birth. However, since the number of reported deaths are smaller nearer the end of the time period in the treated as well as the untreated districts, we restrict our analysis to modeling the hazard during the first 18 days after birth. Hence, we use 18 days as our censoring point instead of 28 days. The frequency distribution of survival information by treatment status is 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 that (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 and that (ii) the observed frequencies exhibit heaping at days multiples of 5 in both samples. The model includes various control variables at the parental level as well as the child level. Summary statistics for these variables distinguished by treatment status can be found in Section S5 of the supplementary material.

As a preliminary to the estimation of formal models, it is informative to examine the non-parametric estimates of the unconditional discrete hazard function distinguished by the treatment status. These are plotted in Figure 3 and are calculated as the ratio of number of failures during the interval divided by the number of babies alive at the beginning of the interval. 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. Notice however that both the denominator and the numerator suffer from rounding errors and so should be interpreted with care.

Empirical Findings: We estimate the model based on the specification outlined in Section 5. 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

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

[21]The estimated effects were very similar across different thresholds.

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 q_1, 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, q_1, and q_2, respectively. We set $\overline{k} = 12$ relying partially on information from the Program for 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.

We start our empirical analysis by testing for the presence of heaping effects ($\mathbf{H}^{\pi_1}$) via the LR test described in Remark 4.2 of Section 4. The result of this exercise can be found in Table 2, which reports the test statistic as well as the corresponding bootstrapped 95% critical value, and clearly suggests that heaping matters in our model. In fact, examining Table 3, which reports the estimated rounding probabilities together with their deviations for a model that accounts for the introduction of the JSY program, confirms the finding that all heaping parameters are significantly different from zero at conventional levels.

Having established that heaping matters from a statistical point of view, we next turn to the actual analysis of the JSY program: as outlined in Section 5, in a first step we would like to rule out that changes in the reporting behavior (as a result of the policy introduction) confound any observable effect of the program. Therefore, we start by testing $\mathbf{H}^{\pi_3}$, which under the null postulates that all deviations $p_1^{(2)}$ through $q_2^{(2)}$ are jointly equal to zero. Examining the second row of Table 2, we find that we cannot reject this null on a 5% significance level. This is in line with the standard errors of $\widehat{p}_1^{(2)}$, $\widehat{p}_2^{(2)}$, $\widehat{q}_1^{(2)}$, and $\widehat{q}_2^{(2)}$ in Table 3, which all exceed the size of the actual parameter estimates. Thus, albeit theoretically plausible, our findings do not support the conjecture that mothers who delivered in treated districts reported differently from those in untreated ones (e.g., because births were better recorded).

In a second step, we now analyze the effect of the JSY program on the baseline parameters $\gamma(0), \ldots, \gamma(17)$. Figures 4(a) and 4(c) display the estimated coefficients of $\exp(\gamma(\cdot))$ and $\exp(\gamma^{(2)}(\cdot))$, respectively, which have been estimated in exponential form. Thus, note that $\exp(\gamma^{(2)}(\cdot)) = 1$ implies that no change has taken place after the introduction of the policy. Pointwise 95% confidence intervals have been plotted around the coefficient estimates.

While it is straightforward to see from Figure 4(a) that we can reject the null that the (stepwise) integrated baseline hazard is zero for each day except days 11 to 15 at a 5% level (recall that $\gamma(\tau) = \ln \int_{\tau}^{\tau+1} \lambda_0(s)ds$), we note that we are only able to reject (again, at a 5% significance level) the null that $\exp(\gamma^{(2)}(\cdot)) = 1$ after day 4.[22] However, given that the majority of deaths occurs within the first 3 days of birth, this finding can at best be interpreted as weak evidence for a reduction in mortality. In fact, turning back to Table 2, we find that the null of $\mathbf{H}^{\gamma_2}$, which states that at least some of the baseline parameters decreased, cannot be rejected at a 5% significance level either.

For comparison reasons, we contrast these results with estimates from a second specification that neglects heaping altogether, see Figures 4(b) and 4(d): while estimates of $\exp(\gamma(\cdot))$ are generally very similar to the ones displayed in Figure 4(a) (a notable exception are days 10 and 15), we point out that there are some differences in the estimates of

[22]Judging by Figure 4(c), it is also clear that the null of the test $\mathbf{H}^{\gamma_1}$, which postulates that no uniform (i.e., over the first 18 days) downward shift of the baseline hazard took place, cannot be rejected at any conventional significance level as the decision rule is based on individual t-tests for each day.

$\exp(\gamma^{(2)}(\cdot))$ particularly around days 5 and 10, which indicate a significant reduction in the model that takes account of the heaping, but not in the one that ignores heaping altogether.

Summarizing the findings of this section, we note that our estimates and test $\mathbf{H}^{\pi_1}$ suggest clear evidence of heaping in the data. By contrast, our test results do not indicate that the introduction of the JSY program led to a significant reduction in mortality over the first 18 days after birth. That is, while some estimated coefficients of $\exp(\gamma^{(2)}(\cdot))$ indicated a significant reduction of mortality for days 5, 7 to 10, and 12 onwards, we were not able to confirm this finding through a joint hypothesis test on all coefficients as we cannot reject the null of all parameters jointly equal to zero at a 5% significance level (cf. $\mathbf{H}^{\gamma_2}$).

However, we stress that our analysis was conducted using only those babies born in districts that were eventually treated, and hence captures the "intention to treat" effect rather than the actual effect of mothers receiving treatment. Thus, it remains to be established whether the latter effect exhibits a similar pattern, too.

7 Conclusions

India has one of the largest neonatal mortality rates in the world. To address this, 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 parameters in the joint presence of heaping and unobserved heterogeneity. Our identification strategy requires neither administrative data nor multiple measurements, but the presence of a correctly reported duration and of some flat segment(s) 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 heaped durations and over-weighting the contribution of non heaped durations. This adjustment ensures consistent estimation of both heaping and baseline hazard parameters, and we establish the asymptotic properties of the maximum likelihood estimator. Moreover, the paper provides various specification tests that allow, among other things, (i) to check for the presence of heaping effects (ii) to compare different heaping specifications (iii) to test for shifts in the baseline hazard.

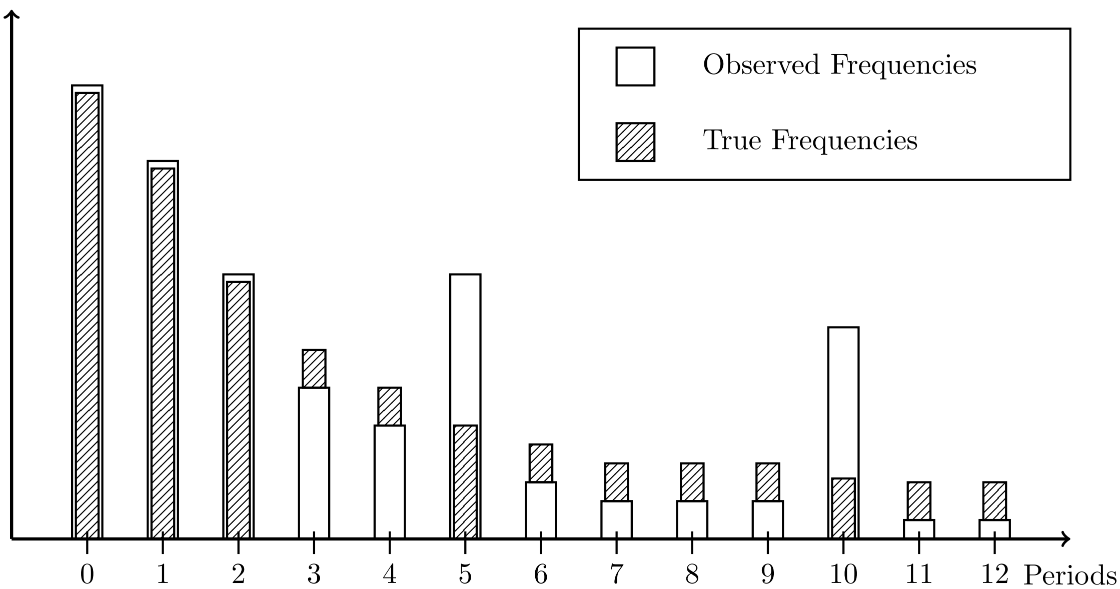
In the supplementary material, we provide an informal discussion of how the idea of our identification strategy can be straightforwardly extended to cover different discrete duration models (e.g., Han and Hausman, 1990; Sueyoshi, 1995) as well as standard ordered choice models. The identification results are therefore not limited to the model used in Section 2, but can also be applied to non-duration data.

Our empirical findings can be summarized as follows: first, a specification test and model estimates indicate that heaping plays an important role in our data. Second, we do not find evidence for an increase in survival probability of babies born in districts that were treated. However, since our analysis only included mothers from districts that were eventually treated, it remains a question for future research whether our findings also hold for the actual effect of mothers receiving treatment.

Finally, another important extension, both of theoretical as well as of empirical relevance, concerns our modeling of the proxy variable for treatment: while some theoretical research has been carried out on studying the effects of measurement error in the treatment variable (e.g. Lewbel, 2007) or more generally in the covariates (Battistin and Chesher, 2014) on treatment effects, it remains unclear how these results translate to cases like ours with (possible) misclassification in the treatment status and heaping in the dependent variable. This is left for future research.

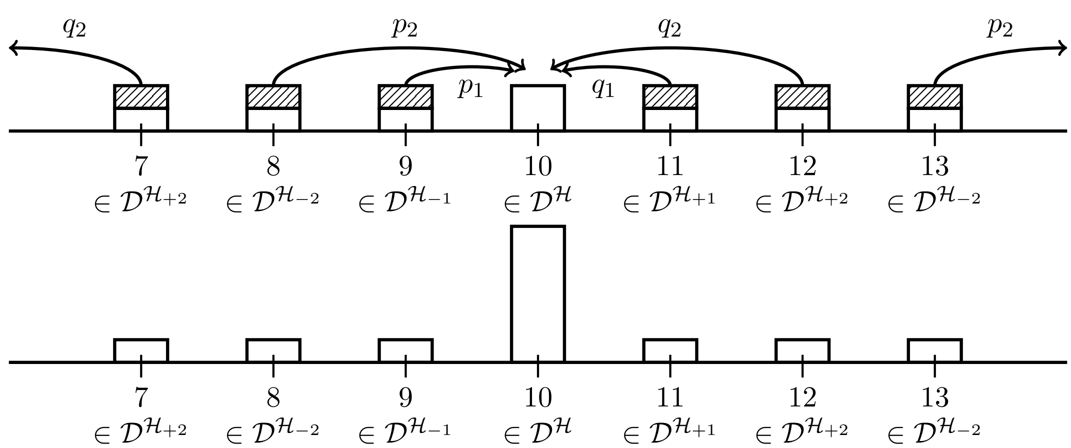
Appendix I

Figure 1: Stylized Failure Frequencies



Note: The figure is a stylized example of a discrete time failure distribution with heaps at time periods 5 and 10. Observed and true frequencies differ from period 3 onwards.

Figure 2: Stylized Heaping Mechanism



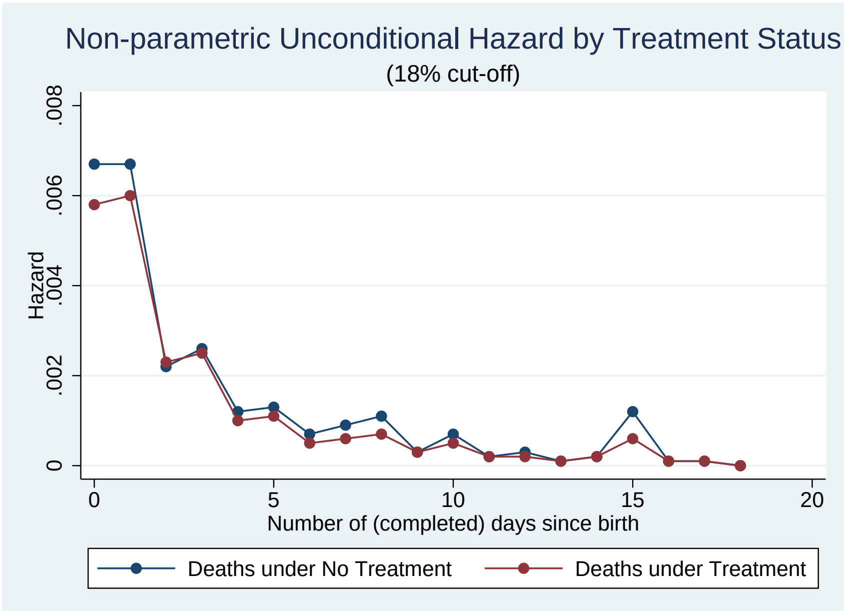
Note: The figure is a stylized example of a potential heaping mechanism for $h^* = 5$, $j = 2$, and $\bar{r} = 2$. The associated heaping probabilities are p_1, p_2, q_1, and q_2.

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

				Untreated Dist.[1]		Treated Dist.[1]	
Days	Freq.	Percent	Cum. Percent	Freq.	Percent	Freq.	Percent
0	1,010	0.64	0.64	793	0.66	217	0.58
1	1,021	0.65	1.30	797	0.67	224	0.60
2	341	0.22	1.51	255	0.21	86	0.23
3	395	0.25	1.77	304	0.25	91	0.24
4	179	0.11	1.88	141	0.12	38	0.10
5	188	0.12	2.00	147	0.12	41	0.11
6	99	0.06	2.06	79	0.07	20	0.05
7	124	0.08	2.14	103	0.09	21	0.06
8	153	0.10	2.24	128	0.11	25	0.07
9	43	0.03	2.27	33	0.03	10	0.03
10	101	0.06	2.33	82	0.07	19	0.05
11	33	0.02	2.35	27	0.02	6	0.02
12	37	0.02	2.38	31	0.03	6	0.02
13	16	0.01	2.39	14	0.01	2	0.01
14	27	0.02	2.40	19	0.02	8	0.02
15	165	0.11	2.51	142	0.12	23	0.06
16	20	0.01	2.52	17	0.01	3	0.01
17	12	0.01	2.53	8	0.01	4	0.01
18	18	0.01	2.54				
19	8	0.01	2.55				
20	65	0.04	2.59				
21	28	0.02	2.61				
22	18	0.01	2.62				
23	5	0.00	2.62				
24	4	0.00	2.62				
25	24	0.02	2.64				
26	6	0.00	2.64				
27	4	0.00	2.64				
Cens. Obs.	152.535	97,36	100	116,169	97.38	36,546	97.74
Total	156,679	100		119,289	100	37,390	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 3: Unconditional Hazards



Note: The discrete hazard was calculated as the ratio of number of failures during the interval divided by the number of babies alive at the beginning of the interval.

Table 2: Specification Tests[1]

	LR-statistic	95% CV[2]
$\mathbf{H}^{\pi_1}$	298.74	288.29
$\mathbf{H}^{\pi_3}$	419.54	619.44
$\mathbf{H}^{\gamma_2}$	507.83	744.47

[1] See Sections 4 and 5 for details of the hypotheses that are being tested.

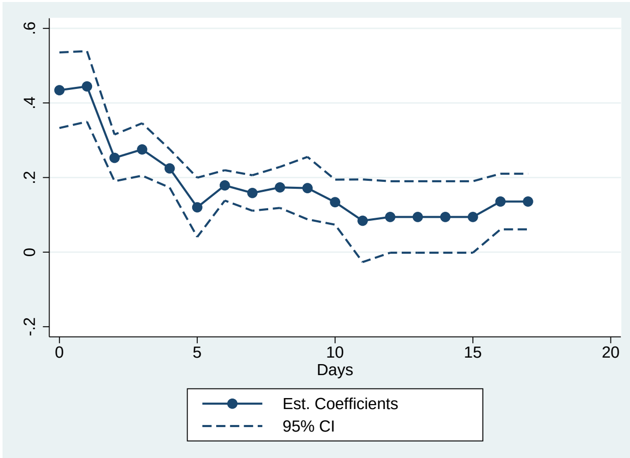
[2] Based on Empirical Bootstrap Distribution with 200 replications.

Table 3: Estimated 'Heaping' Parameters

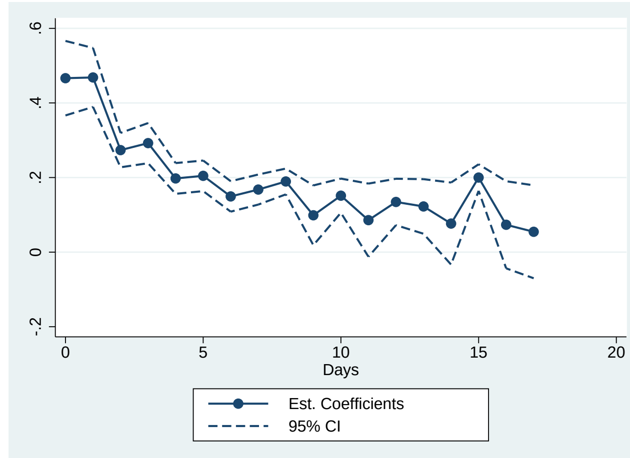
Model with Heaping[1]		
	Coeff.Est.	Bootstrapped S.E.[2]
Heaping Probabilities		
p_1	0.577	0.269
p_2	0.453	0.166
q_1	0.480	0.200
q_2	0.521	0.157
Deviations After Treatment		
$p_1^{(2)}$	− 0.011	0.386
$p_2^{(2)}$	− 0.232	0.658
$q_1^{(2)}$	− 0.207	0.552
$q_2^{(2)}$	− 0.137	0.619

[1] Model allows for small heaps at days 5 and 10 with associated probabilities p_1 and q_1, and a large heap at day 15 with associated probabilities p_1, p_2, q_1, and q_2.

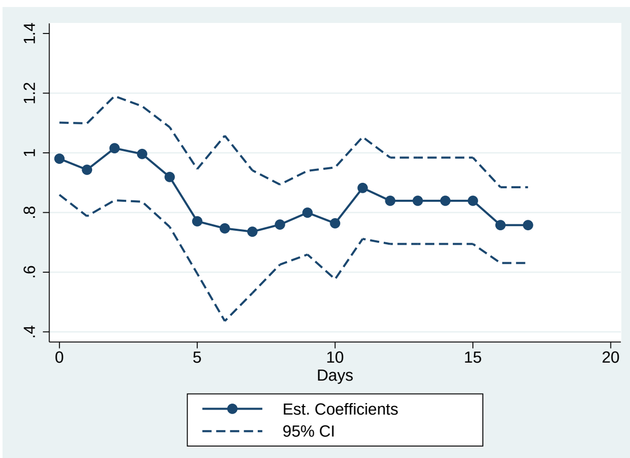
[2] Bootstrapped standard errors with 200 replications.



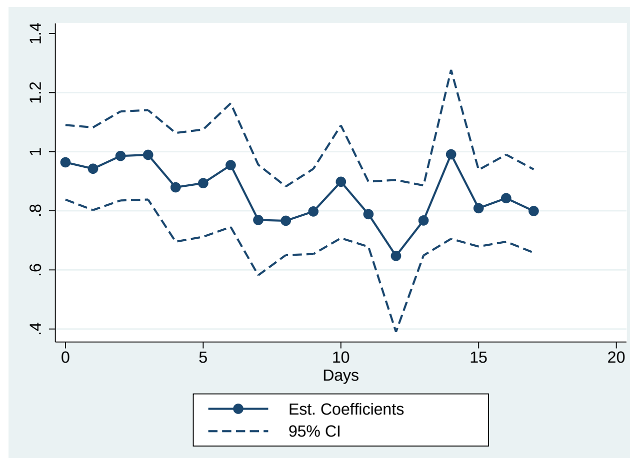
(a) Heaping Model: $\exp(\hat{\gamma})$ by days



(b) Non-Heaping Model: $\exp(\hat{\gamma})$ by days



(c) Heaping Model: $\exp(\hat{\gamma}_2)$ by days



(d) Non-Heaping Model: $\exp(\hat{\gamma}_2)$ by days

Figure 4: Estimated Baseline Hazard and Deviation Coefficients (with pointwise 95% CI)

Appendix II

Proof of Proposition 1: In the following, suppose that the flat segment lies around the last heap $\overline{j} \cdot h^*$ and that $\overline{k} = \overline{j} \cdot h^* - \overline{r} - 1$ and $\overline{\tau} = \overline{j} \cdot h^* + \overline{r} + 1$. Note that setting the flat region around the last heap $\overline{j} \cdot h^*$ comes at no loss of generality since the main steps of the proof remain identical if $j < \overline{j}$ (see Remark A1 below). We set $\overline{r} = 1$, which implies that $\overline{j} \cdot h^* = \overline{k} + 2$. The extension to $\overline{r} > 1$ will be outlined subsequently. Finally, without loss of generality, assume that z_i is a scalar.

Define for any time period $\tau < \overline{\tau}$

$$H_0(\tau) = \sum_{s=0}^{\tau} \exp\left(\gamma(s)\right)$$

as the discrete cumulative baseline hazard.

First of all, notice that $\overline{k}$ is correctly observed by H(iv) 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,

$$\begin{aligned}
\Pr\left(\tau_i \geq \overline{k}+1|z_i, \theta\right) &= S_i\left(\overline{k}+1|z_i, \theta\right) \\
&= \int_0^\infty S_i\left(\overline{k}+1|z_i, v, \underline{\theta}\right) g(v;\sigma)dv \\
&= \int_0^\infty \exp\left(-vH_0(\overline{k})\exp\left(z_i\beta\right)\right) g(v;\sigma)dv \\
&= \left(1+\sigma\left(H_0(\overline{k})\exp\left(z_i\beta\right)\right)\right)^{-\sigma^{-1}}.
\end{aligned}$$

Since the covariates are time invariant and independent of unobserved heterogeneity, set $z_i = 0$ to obtain $S_i\left(\overline{k}+1|z_i=0, \theta\right)$. 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:

$$A(H_0(\overline{k})) = \int_0^\infty \exp\left(-vH_0(\overline{k})\right) g(v;\sigma)dv.$$

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(ii) 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 $\gamma(\overline{k}) = \overline{\gamma}$. By assumption H(iii), this implies that also $H_0(\overline{k}+1) = H_0(\overline{k}+2)$ and $\gamma(\overline{k}+1) = \gamma(\overline{k}+2) = \overline{\gamma}$ 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 is identified by standard arguments under Assumption Z. 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 $\Pr\left(\tau_i = \overline{k}+3|z_i=0,\theta\right)$ 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, $\gamma(\overline{k}+3) = \overline{\overline{\gamma}}$ can be identified from $H_0(\overline{k}+3) - H_0(\overline{k}+2)$.[23]

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

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

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.[24] Now,

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

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 $H_0(h^*-1)$.

[23] Note that $\Pr\left(t_i \geq \overline{k}+4|z_i,\theta\right) = \Pr\left(\tau_i \geq \overline{k}+4|z_i,\theta\right) = S_i\left(\overline{k}+4|z_i,\theta\right)$ as $\overline{k}+4=\overline{\tau}$, and $\overline{\tau} \in \mathcal{D}^{\mathcal{T}}$ (if this is not the case, $(\overline{k}+4) \in \mathcal{D}^{\mathcal{H}_{-l}}$, $l=\overline{r}$, and the same result applies).

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

Next, recalling

$$\Pr\left(t_i = h^* - 1 | z_i = 0, \theta\right) = (1 - p_1) \Pr\left(\tau_i = h^* - 1 | z_i = 0, \theta\right),$$

$$\Pr\left(t_i = h^* + 1 | z_i = 0, \theta\right) = (1 - q_1) \Pr\left(\tau_i = h^* + 1 | z_i = 0, \theta\right),$$

and

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

it follows that

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

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 $\gamma(h^* + 1)$,

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

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

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

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

where we used the fact that γ is constant after $\overline{j} \cdot h^*$. Thus, since $\overline{r}$ is known,

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

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

$$\Pr\left(t_i = \overline{j} \cdot h^* - l | z_i = 0, \theta\right) = (1 - p_l) \Pr\left(\tau_i = \overline{j} \cdot h^* - l | z_i = 0, \theta\right)$$

and

$$\Pr\left(t_i = \overline{j} \cdot h^* + l | z_i = 0, \theta\right) = (1 - q_l) \Pr\left(\tau_i = \overline{j} \cdot h^* + l | z_i = 0, \theta\right).$$

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 $\overline{r} = 1$. ■

Remark A1: If the flat segment does not contain the last heap $\overline{j} \cdot h^*$, but a heap with $j < \overline{j}$, the structure of the proof is as follows: First, as in the proof of Proposition 1, identify p_l and q_l, $l = 1, \ldots, \overline{r}$, as well as $\overline{\gamma}$ and $\overline{\overline{\gamma}}$. Knowing the heaping probability parameters, one can proceed to identify all $\gamma(k)$ parameters prior to $\overline{k}$. Subsequently, all $\gamma(k)$ parameters with $j \cdot h^* + \overline{r} < k < \overline{\tau}$ can be identified.

Proof of Theorem 2:
(i) Given Assumption D, by the uniform law of large numbers 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 Assumptions 2-6 hold. Note that, given Assumption U,

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

and from the definition of $\phi_i(\cdot)$ in (I)-(IV), it is immediate to see that $l_N(\theta)$ has well defined left and right derivatives for $\theta \in \Psi^+$, with $\Psi^+ = \Psi \cap C\left(\theta^{\dagger}, \varepsilon\right)$ and $C\left(\theta^{\dagger}, \varepsilon\right)$ 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 that 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(ii), the boundary issue may arise when some p_l and/or q_l are zero for $l = 1, ..., \overline{r}$. Hence, $\left(\Theta - \theta^{\dagger}\right)$ is locally equal to Ψ, which is a convex cone in $\mathbb{R}^{K_\beta + K_\gamma} \times \mathbb{R}_+ \times [0,1)^{2\overline{r}}$ and Assumptions 5 and 6 in A99 hold.
(iii) In this case $\theta^{\dagger}$ is not on the boundary, and so

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

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

References

Augustin, T. and J. Wolff (2004). A Bias Analysis of Weibull Models under Heaped Data. *Statistical Papers*, 45, 211-229.

Abbring, J.H. and G. Van Den Berg (2007). The Unobserved Heterogeneity Distribution in Duration Analysis. *Biometrika*, 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.

Battistin, E. and A. Chesher (2014). Treatment Effect Estimation with Covariate Measurement Error *Journal of Econometrics*, 178, 707-715.

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*, 30, 353-376.

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.

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. Working Paper. URL: `http://econweb.umd.edu/~ham/seam_bias_2014_03_17jh_ls_accept.pdf`

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 (2014). Estimating a Semiparametric Duration Model without Specifying Heterogeneity. *Journal of Econometrics,* 178, 114-131.

Heitjan, D.F. and D.B. Rubin (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.

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.

Lewbel, A. (2007). Endogenous Selection or Treatment Model Estimation. *Econometrica*, 141, 777-806.

Lim, S., Dandona, L., Hoisington, J., James, S., Hogan, M., and E. Gakidou, (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 A. Kanchi (2013). *Women's Work, Maternity and Public Policy in India*, TATA Institute of Social Sciences, Hyderabad, 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).

Petoussis, K., R.D. Gill and C. Zeelenberg (1997). Statistical Analysis of Heaped Duration Data. Vrije University. Working Paper. URL: `http://dspace.library.uu.nl/bitstream/handle/1874/2089/cos.pdf?sequence=1`

Powell-Jackson, T., Mazumdar, S. and A. Mills (2015). Financial Incentives in Health: New Evidence from India's Janani Suraksha Yojana. *Journal of Health Economics*, 43, 154-169.

Ridder, G., T.M. Woutersen (2003). The Singularity of the Information Matrix of the Mixed Proportional Hazard Model. *Econometrica*, 71, 1579-1589.

Roy, S.S. and Others (2013). Improved Neonatal Survival After Participatory Learning and Action With Women's Groups: A Prospective Study in Rural Eastern India. *Bulletin World Health Organization*, 91, 426-433B.

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

Sueyoshi, G. T. (1995). A Class of Binary Response Models for Grouped Duration Data, *Journal of Applied Econometrics*, 10, 411-431.

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

Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.

Wolff, J. and T. Augustin (2003). Heaping and its Consequences for Duration Analysis: a Simulation Study. *Allgemeines Statistisches Archiv*, 87, 59-86.