White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/244746/

Version: Accepted Version

Article:
Taylor, C. C., Di Marzio, M., Fensore, S. et al. (Accepted: 2026) Recovering event-time distributions under multiresolution time rounding: the case of crime data. Journal of the Royal Statistical Society Series C: Applied Statistics. ISSN: 0035-9254 (In Press)

https://doi.org/10.1093/jrsssc/qlag051

This is an author produced version of an article accepted for publication in Journal of the Royal Statistical Society Series C: Applied Statistics, made available via the University of Leeds Research Outputs Policy under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Journal of the Royal Statistical Society Series C, 2026, 1–18
doi: DOI to be added during production
Published: Date added during production
Paper

Recovering event-time distributions under multiresolution time rounding: the case of crime data

Charles C. Taylor,[1,*] Marco Di Marzio,[2] Stefania Fensore[2] and Chiara Passamonti[3]

[1]School of Mathematics, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK, [2]Department of Socio-Economic, Management and Statistical Studies, University of Chieti-Pescara, Viale Pindaro 42, Pescara, 65129, Italy and [3]Department of Mathematics, University of Luxembourg, 2 Avenue de l'Université, Esch-sur-Alzette, L-4365, Luxembourg

*Corresponding author: Charles C. Taylor, charles@maths.leeds.ac.uk

FOR PUBLISHER ONLY Received on Month 2025; revised on Month 2025; accepted on Month 2025

Abstract

Reported time-of-day data are often affected by measurement error due to rounding and digit preference, producing artificial temporal spikes and distorted diurnal patterns. These artifacts arise from heterogeneous reporting practices, with incidents recorded at varying resolutions depending on context and event type. We address the problem of recovering the underlying distribution of event times when only multiresolution rounded observations are available. Our approach estimates rounding probabilities through a constrained least-squares decomposition of the empirical distribution of "minutes past the hour," using these estimates as an intermediate step for penalized inversion over the 24-hour cycle. Because clock-time is circular, we explicitly ensure circular coherence both within-hour and across-hour scales. Motivated by Boston crime data, the method reveals substantial heterogeneity in reporting precision across offense types: aggravated assaults are mostly recorded at the minute level, whereas auto thefts show extensive coarse rounding. The method is broadly applicable to settings where reported times undergo multi-resolution rounding, including medical, survey, and administrative data.

Key words: Boston crime data, Circular observations, Digit preference, Measurement errors, Multiresolution rounding

1. Introduction

The analysis of clock-time data in fields such as criminology, epidemiology, and public health is commonly used to study daily patterns, identify periods of elevated risk, and evaluate interventions. However, the accuracy of clock-time data is critically dependent on the precision with which times are recorded or known. Very often these data are fundamentally compromised by measurement errors, most notably rounding and digit preference, i.e. the human tendency to gravitate toward or away from specific numbers. As a consequence, these reporting artifacts affect the estimate of the distribution of event times and, if unaddressed, can lead to misleading conclusions in both descriptive and inferential analyses. Such errors stem from various sources: for example, in administrative and survey contexts, much of the discrepancy in reported times arises from the habitual practices of data recorders and the fixed conventions of data-collection systems, as pointed out by Groves et al. (2011). In criminological records, discrepancies in recorded incident times often

© The Author 2026. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

reflect idiosyncratic reporting practices among officers or variations in departmental protocols, rather than true event dynamics (see Klinger and Bridges (1997)). More broadly, errors arise from individual cognitive factors, including memory limitations, imprecise information, or intentional approximation. For instance, Bopp and Feh (2008) demonstrated that preferences for terminal digits in self-reported height vary by spoken language, underscoring the role of cultural and cognitive factors. Other studies have focused on self-reported anthropometric data (height, weight, blood pressure), price rounding, etc. (see Hand et al. (2000) and Hand et al. (2001)).

In criminology studies, correctly understanding when a crime occurs informs theories of opportunity and routine activities, guides the deployment of patrol resources, and provides evidence for evaluating prevention interventions (see, respectively, Cohen et al. (1979), Sherman and Weisburd (1995), Farrington et al. (2002)). In the case of crime recordings, uncorrected data registered by the police, such as the Boston one analyzed in this work, show large apparent peaks in criminal activity at midnight and noon, forcing a naive analyst to allocate additional patrol resources to these hours, when in fact the peaks are purely artifactual. Conversely, genuine temporal patterns such as the early-morning crime minimum predicted by routine activity theory may be obscured by the spreading effect of coarse time rounding. The stakes extend beyond resource allocation. Legal proceedings may turn on the precise timing of incidents. Program evaluations comparing crime rates before and after an intervention rely on accurate temporal distributions. Predictive policing algorithms trained on rounded data will learn and perpetuate the reporting artifacts rather than genuine crime patterns.

Similar rounding and digit preference effects arise in many applied settings beyond criminology. Recent empirical work shows that rounding is not merely a legacy data-quality problem, but a modern source of systematic distortion in contemporary survey and administrative data. Tail and center rounding affects reported probabilistic expectations in the Health and Retirement Study, as pointed out by Giustinelli et al. (2022); travel surveys exhibit rounding in reported departure times, with differences by trip purpose and transport mode (see Sato and Maruyama (2020)). Sanko and Iriguchi (2022) shows that comparisons with times recorded by an objective third party indicate that self-reported times may be substantially rounded. Arrival times in emergency departments are also often recorded with heterogeneous precision depending on urgency and reporting protocols, while traffic accident times are frequently approximated and rounded to salient clock values.

Most methodological work on rounding concerns linear variables or scalar survey responses. Circular time-of-day data raise additional difficulties because rounding interacts with the topology of the sample space, especially near the boundary between the end and beginning of the day. Recent work proposed by Jammalamadaka et al. (2024) on testing uniformity on the circle when observations are rounded highlights the importance of rounding in circular settings. Relatedly, statistical methods for measurement error are well-developed for continuous variables subject to additive noise in the so-called errors-in-variables context. Deconvolution kernel density estimation, introduced by Stefanski and Carroll (1990) and its extensions to circular data, developed by Di Marzio et al. (2022) and Di Marzio et al. (2023), can recover underlying distributions when observations are perturbed by independent random errors. However, time rounding is fundamentally different: it is a many-to-one transformation where true times within specific intervals all map to the same recorded value. For example, an actual time of 12:07 might be recorded, according to different rounding resolutions, as either 12:00, 12:05, or 12:07. In addition, an actual time of 12:00 will not generally be rounded so the "errors" between the recorded times and the actual times are not identically distributed. This discrete, resolution-dependent structure cannot be captured by convolution models.

A unified framework that models the probabilities of rounding to different discrete resolutions within the circular setting is currently lacking. The need to fill this gap is motivated by data such as the Boston crime incident reports, which reveal pronounced heterogeneity in rounding behavior. For example, auto thefts show strong hourly rounding, likely due to uncertain timing, whereas aggravated assaults are recorded with high minute-level precision. This indicates that rounding is not a uniform noise process but a context-dependent reporting behaviour, varying by event type and circumstance.

In this paper we address the problem of estimating true event-time distributions from rounded clock-time data. We begin by estimating multiresolution rounding probabilities (minute, 5-minute, 10-minute, etc.) through a constrained least-squares decomposition of the empirical distribution of recorded minutes past the hour. These probabilities are then incorporated into a matrix representation of the rounding

mechanism, which allows us to express the relationship between actual and recorded times in a unified algebraic form. Building on this structure, we extend the approach to the full 24-hour cycle by constructing hour-specific rounding matrices that account for potential diurnal variation in reporting accuracy. Finally, we recover the underlying time-of-day distribution via a penalized inversion procedure that produces smooth and interpretable estimates and we then demonstrate its effectiveness through an application to Boston crime data and a simulation study. Both within-hour data and full-hour measurements possess an inherently circular structure, and our methodology is explicitly designed to respect this property, ensuring that the resulting estimates remain coherent on the circular time scale.

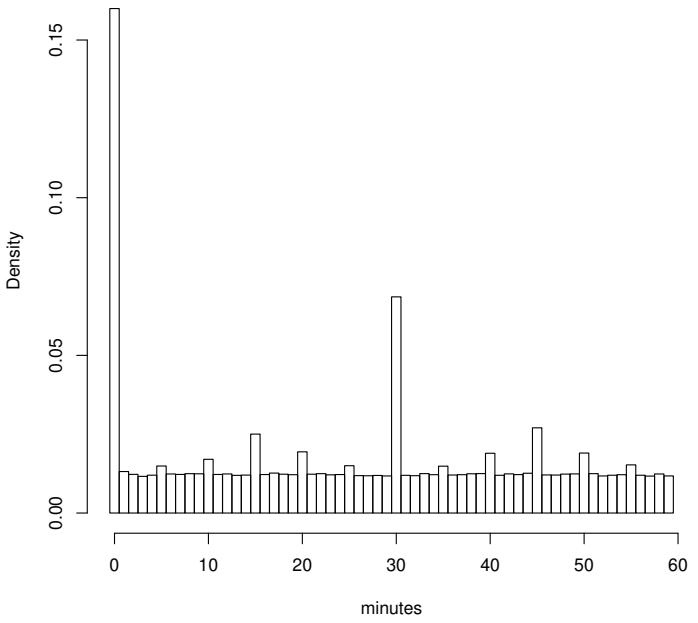

Fig. 1: Density scaled histogram of the recorded time (minutes past the hour) of 319,073 reported crimes in Boston.

The paper is organized as follows. Section 2 describes the common phenomenon of multiresolution time rounding and details the constrained least-squares estimation of rounding probabilities. The results are compared across different crime types. Section 3 extends the model to the daily cycle, and a discussion of the smoothing parameter selection is also included. In Section 4 we provide a brief simulation experiment, and concluding remarks with further considerations are offered in Section 5.

2. Time recording preferences

2.1. Boston Crime Data

Our empirical analysis uses incident-level data from the Boston Police Department (PD) covering 15 June, 2015 to 3 September, 2018, comprising 319,073 reported incidents across 36 offence categories. These data are publicly available through the city's open data portal at the link https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system, and represent all incidents recorded in the

department's Records Management System (RMS), excluding only those cases where public disclosure would compromise ongoing investigations or victim privacy. Boston PD, like most modern urban police departments, uses a Computer-Aided Dispatch (CAD) system for real-time incident response. When a 911 call is received or an officer initiates an incident, the CAD system records the incident time automatically. However, many incidents are not reported contemporaneously but are instead documented hours or days after occurrence through follow-up reports. In these cases, the "incident time" field represents the officer's best estimate based on victim/witness statements, surveillance footage time stamps, or simply the midpoint of a known time window. Officers are instructed to record the most accurate time possible, even when there is much uncertainty.

Figure 1 shows a density histogram of recorded "minutes past the hour". At present, we do not consider the hour of the time of day, only the number of minutes after the o'clock. The picture shows an underlying flat distribution but with a major peak at 0, a large peak at 30, smaller peaks at 15 and 45, and minor peaks at 5, 10, 20, 25 etc. This is linked to the commonly observed phenomenon, in which recorded times are usually rounded to the nearest minute, but sometimes rounded to the nearest 5 minutes, nearest quarter of an hour, or even to the nearest hour. Such rounding behavior is heterogeneous and context-dependent, arising from the reporter's habits, memory limitations, or uncertainty about the precise time. This heterogeneity introduces a structured bias into the observed distribution, which needs to be accounted for to recover the true distribution of the event times.

We consider the task of estimating the distribution of actual times, given (a) the recorded times, and (b) a "rounding preference"probability, which must first be estimated. To help interpretation, we will use minutes past the hour as our unit of measurement, rather than the angle of the time measured in degrees or radians, but continue to recognize the circular aspect of the data. This means that while we use linear labels, our calculations account for the fact that the scale wraps around (e.g., minute 59 is adjacent to minute 0). For example, an actual time of 58 minutes (past the hour) could be recorded as 0, since on the circular timeline these two points are neighbours.

2.2. Notation and model

We formalize the rounding mechanism by explicitly modelling the relationship between true event times, rounding resolutions, and recorded times. Let the circular random variable T denote the actual time at which an event occurs. According to the convention introduced earlier, it takes values within $(0, 60]$ with $0 \equiv 60$. We assume that a probability density function f_T exists. Time is measured in minutes, so rounding of t always occurs (at least to the nearest minute). Let K be a discrete random variable describing the rounding resolution, taking values in the set

$$\mathcal{K} = \{1, 5, 10, 15, 30, 60\}.$$

For each $k \in \mathcal{K}$, define the *rounding set*

$$R_k = \{\, i \in \mathbb{N} \cap (0, 60] \mid i \mod k = 0\}, \quad \text{so that} \quad |R_k| = 60/k.$$

Example 1 *For $k = 15$ we have $R_{15} = \{15, 30, 45, 60\}$, where only one of the labels 0 or 60 is included due to circularity, and we confirm $|R_{15}| = 60/15 = 4$.*

Given an actual time t and a fixed rounding resolution k, define the deterministic recorded value

$$y_k(t) := \mathrm{proj}_{R_k}(t) = \arg\min_{x \in R_k} d(t, x), \tag{1}$$

where $d(w, z)$ is a circular distance between times w and z. Thus $y_k(t)$ is the value that would be recorded if the actual time were t and the chosen rounding resolution were k. If both the actual time T and the rounding

resolution K are random, the rounded minute is a random variable Y, where

$$Y = y_K(T).$$

Consequently, conditional on $K = k$ and $T = t$, the event $Y = y$ is equivalent to the deterministic statement $y_k(t) = y$.

Example 2 *We provide some examples for various t and k: $y_{10}(23.1) = 20,\ y_{15}(39.7) = 45,\ y_{60}(11.5) = 60$, where $60 \equiv 0$ on the circular time scale.*

For a fixed resolution k, the conditional probability of obtaining minute y when the event occurs in interval $\mathcal{I}_j = [j - 1/2, j + 1/2)$ is

$$\mathsf{P}(Y = y \mid K = k,\ T \in \mathcal{I}_j) = \mathsf{P}\{y_k(T) = y \mid T \in \mathcal{I}_j\}, \quad j = 1, 2, \ldots, 60.$$

The value of this probability depends on both the structural properties of the rounding sets R_k and the location of the actual time within the interval $\mathcal{I}_j$ relative to y.

Example 3 *We provide three illustrative cases to reflect distinct probabilistic outcomes.*

(i) $\mathsf{P}(Y = 60 \mid K = k,\ T \in [59.5, 0.5)) = 1$ *for all $k \in \mathcal{K}$, since $60 \in R_k$ for every allowed resolution.*
(ii) $\mathsf{P}(Y = 1 \mid K = k,\ T \in [2.5, 3.5)) = 0$ *for the resolutions considered here, since times in this interval do not round to 1.*
(iii) $\mathsf{P}(Y = 15 \mid K = k,\ T \in [16.5, 17.5)) > 0$ *for resolutions such as $k = 5$ or $k = 15$, because times in this interval may be mapped to the attraction point 15.*

A given recorded value y may correspond to several combinations of rounding resolution and actual time satisfying $y_k(t) = y$. Hence the mapping from (k, t) to y is generally *many-to-one*. The marginal probability of an observed rounded minute is obtained by summing over rounding resolutions and integrating over all actual times that would be recorded as y. To obtain the finite-dimensional representation in the second line below, the continuous circle is partitioned into the one-minute intervals $\mathcal{I}_j$, and $y_k(j)$ denotes the rounded value assigned to interval $\mathcal{I}_j$ under the chosen discretization, with boundary cases treated according to the rounding tie convention.

Without assuming independence between K and T, we have

$$\begin{aligned} \mathsf{P}(Y = y) &= \sum_{\{k \in \mathcal{K}\}} \int_0^{60} \mathbf{1}(y_k(t) = y)\, \mathsf{P}(K = k \mid T = t)\, f_T(t)\, dt \\ &= \sum_{\{k \in \mathcal{K}\}} \sum_{j=1}^{60} \mathbf{1}(y_k(j) = y) \int_{\mathcal{I}_j} \mathsf{P}(K = k \mid T = t)\, f_T(t)\, dt. \end{aligned} \tag{2}$$

Define the joint probability that an event occurs in the interval $\mathcal{I}_j$ and is rounded according to resolution k as

$$p_{kj} := \mathsf{P}(K = k,\, T \in \mathcal{I}_j) = \int_{\mathcal{I}_j} \mathsf{P}(K = k \mid T = t)\, f_T(t)\, dt. \tag{3}$$

By construction, $\sum_{k \in \mathcal{K}} p_{kj} = \mathsf{P}(T \in \mathcal{I}_j)$ and $\sum_{j=1}^{60} \sum_{k \in \mathcal{K}} p_{kj} = 1$. Using the above definition, Equation (2) can be written as

$$\begin{aligned} \mathsf{P}(Y = y) &= \sum_{\{k \in \mathcal{K}\}} \sum_{j=1}^{60} p_{kj}\, \mathbf{1}(y_k(j) = y) \\ &= \sum_{\{(k,j):\, y_k(j) = y\}} p_{kj}. \end{aligned} \tag{4}$$

Example 4 *The probability of recording the minute $y = 5$ is given by*

$$\mathsf{P}(Y = 5) = p_{15} + \sum_{j=3}^{7} p_{5j},$$

since the only rounding resolutions that can produce the recorded value $y = 5$ are $k = 1$ and $k = 5$. Specifically, the first term corresponds to events occurring in the interval $\mathcal{I}_5 = [4.5, 5.5)$ and rounded to the nearest minute. The second term aggregates events occurring in the intervals $\mathcal{I}_3, \ldots, \mathcal{I}_7$ that, when rounded to the nearest 5 minutes, yield the recorded value $y = 5$.

2.3. Estimation

We now turn from the general representation in terms of the joint probabilities p_{kj} to the estimation of the marginal rounding weights $p_k = \mathsf{P}(K = k)$. The first-stage decomposition is based on the following local identifying assumption: in the absence of rounding, actual event times are approximately uniformly distributed within each hour. This assumption is used only as a within-hour baseline for separating rounding artefacts from the latent event-time distribution. It is a behavioural approximation, motivated by the fact that the actual time (minutes past the hour) at which a crime takes place is unlikely to be tied to the precise clock time. A possible exception to this may be where an intruder uses the chiming of a clock to mask the noise of a break-in. In contrast, the rounding patterns studied here arise from reporting behaviour, recall uncertainty, administrative conventions, and operational protocols. Specifically, we assume

$$\int_{\mathcal{I}_j} f_T(t)\, dt = \frac{1}{60}, \qquad j = 1, \ldots, 60.$$

This assumption allows the effect of a rounding resolution k to be expressed only through the length of the interval of true times mapped to each recorded minute. Without such a restriction, the observed minute-past-the-hour distribution would generally not identify both the latent within-hour distribution and the rounding probabilities. We do not claim that this approximation is exact for every offence type or every period of the day; rather, it provides a parsimonious local baseline.

Combining within-hour uniformity with the working independence assumption between K and the actual within-hour time gives $p_{kj} = p_k/60$. For a fixed resolution k, a recorded attraction point $i \in R_k$ collects a total interval length k of actual times; hence its contribution to the probability of observing minute i is $(k/60)p_k$. Let $\boldsymbol{p} = (p_1, p_5, p_{10}, p_{15}, p_{30}, p_{60})^\top$ denote the vector of marginal rounding probabilities, where $p_k = \mathsf{P}(K = k)$, $p_k \geq 0$, and $\sum_{k \in \mathcal{K}} p_k = 1$. Thus, the probability of observing minute i can be written as

$$\pi_i(\boldsymbol{p}) = \mathsf{P}(Y = i) = \sum_{k \in \mathcal{K}} a_{ik} p_k, \qquad a_{ik} = \frac{k}{60}\, \mathbf{1}(i \in R_k). \tag{5}$$

Let N_i be the number of observations recorded at minute i past the hour and let $f_i = N_i/n$, where $n = \sum_i N_i$. Our main estimator is a descriptive constrained least-squares, or minimum-distance, fit of the empirical minute-past-the-hour distribution to the model-implied distribution:

$$\hat{\boldsymbol{p}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{p} \in \Delta} \sum_{i=1}^{60} \left[f_i - \pi_i(\boldsymbol{p}) \right]^2, \qquad \text{where} \quad \Delta = \left\{ \boldsymbol{p} : p_k \geq 0,\ \sum_{k \in \mathcal{K}} p_k = 1 \right\}. \tag{6}$$

This criterion should not be interpreted as an ordinary regression model with independent Gaussian errors. We use the constrained least-squares criterion only as a descriptive calibration device for decomposing the marginal distribution of recorded minutes into interpretable rounding components.

For comparison, under an independent-event sampling approximation, a natural likelihood benchmark is obtained by treating the count vector as multinomial,

$$(N_1, \ldots, N_{60}) \sim \text{Multinomial}\{n; \pi_1(\boldsymbol{p}), \ldots, \pi_{60}(\boldsymbol{p})\},$$

which leads to the constrained maximum-likelihood estimator

$$\hat{\boldsymbol{p}}_{\text{ML}} = \arg\max_{\boldsymbol{p} \in \Delta} \sum_{i=1}^{60} N_i \log \pi_i(\boldsymbol{p}). \tag{7}$$

Thus $\hat{\boldsymbol{p}}_{\text{ML}}$ estimates the same vector of marginal rounding probabilities as $\hat{\boldsymbol{p}}_{\text{LS}}$, but using the multinomial likelihood criterion. In the present global minute-past-the-hour model, the two criteria lead to the same estimator when the implied rounding weights lie in the interior of the simplex. The rounding-mixture model induces only six distinct fitted probability levels among the 60 possible minutes past the hour. Specifically, using the minute labels $1, \ldots, 60$, let

$$\mathcal{G} = \{G_1, G_5, G_{10}, G_{15}, G_{30}, G_{60}\},$$

where G_1 contains the minutes that are not multiples of five, $G_5 = \{5, 25, 35, 55\}$, $G_{10} = \{10, 20, 40, 50\}$, $G_{15} = \{15, 45\}$, $G_{30} = \{30\}$, and $G_{60} = \{60\}$, with $60 \equiv 0$. Each minute belongs to exactly one group in $\mathcal{G}$. Minutes in the same group receive the same combination of rounding contributions and therefore have the same model-implied probability $\pi_i(\boldsymbol{p})$. For $g \in \{1, 5, 10, 15, 30, 60\}$ and $i \in G_g$, denote this common probability by θ_g.

For the least-squares criterion, since $\pi_i(\boldsymbol{p}) = \theta_g$ for all $i \in G_g$, the objective (6) can be written groupwise as $\sum_g \sum_{i \in G_g} (f_i - \theta_g)^2$. Thus, for each group, we have

$$\hat{\theta}_g^{\text{LS}} = \frac{1}{|G_g|} \sum_{i \in G_g} f_i.$$

For the multinomial likelihood, let $N_g = \sum_{i \in G_g} N_i$ denote the total count in group G_g. Since all cells in the same group have fitted probability θ_g, the log-likelihood (7) can be written as $\sum_g N_g \log \theta_g$, with the constraint $\sum_g |G_g| \theta_g = 1$. The Lagrange equations give

$$\hat{\theta}_g^{\text{ML}} = \frac{N_g}{n|G_g|} = \frac{1}{|G_g|} \sum_{i \in G_g} f_i.$$

Thus, for this grouped rounding model, constrained least squares and multinomial maximum likelihood produce the same fitted minute probabilities. The mapping from rounding weights to group probabilities is triangular: each successive group probability adds one new rounding weight to those already appearing in the previous groups. Therefore, once the group probabilities have been estimated, the rounding weights are recovered sequentially and uniquely. Since constrained least squares and multinomial likelihood give the same estimated group probabilities, they also give the same rounding weights, provided that these weights lie in the simplex.

Applying the constrained estimation procedure to the Boston crime data yields the rounding-weight estimates reported in Table 1. We see that all estimated rounding weights are positive; hence the solution lies in the interior of the simplex. Nearly three-quarters of events are precise to the minute, while about 9% are recorded only to the nearest hour. The high spike in $60 \equiv 0$ minutes past the hour can arise from any of the above rounding resolutions, since $60 \in R_k$ for all $k \in \mathcal{K}$. Events recorded exactly at the hour are therefore a pooling of true times from different intervals, depending on the rounding resolution. For example, if rounding is to the nearest minute, a true time in the circular interval $[59.5, 0.5)$ is recorded as 0; if rounding is to the nearest 5 minutes, true times in the wider interval $[57.5, 2.5)$ can be rounded to 0; for $k = 15, 30, 60$ the interval of true times that maps to 0 becomes progressively wider.

Rounding resolution k	Estimated rounding weight $\hat{p}_k$
1	0.73
5	0.03
10	0.02
15	0.04
30	0.08
60	0.09

Table 1. Estimated rounding weights for the Boston crime data. Values are rounded to two decimal places. For the global minute-past-the-hour model, the constrained least-squares and multinomial maximum-likelihood estimators coincide analytically because the solution lies in the interior of the simplex.

2.4. An alternative construction

Let $\mathbf{f} = (\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_{60})^\top$ be the probability distribution of rounded times, where f_i represents the probability of recording i minutes past the hour. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_{60})^\top$ be the vector representing the probability distribution of actual event times, where $x_j = \mathsf{P}(T \in \mathcal{I}_j)$. Let $\boldsymbol{P}$ be the 60×60 matrix representing the conditional rounding probabilities, where the element P_{ij} gives the probability that a time is recorded as i given that the actual time falls in the interval $\mathcal{I}_j$:

$$P_{ij} = \mathsf{P}(Y = i \mid T \in \mathcal{I}_j) \qquad i = 1, 2, \ldots, 60 \qquad j = 1, 2, \ldots, 60.$$

The matrix $\boldsymbol{P}$ transforms the probability distribution of actual times $\boldsymbol{x}$ into the probability distribution of rounded times $\mathbf{f}$ through the relationship:

$$\mathbf{f} = \boldsymbol{P}\boldsymbol{x}. \tag{8}$$

This identity follows by expressing the rounded-time probabilities as

$$\begin{aligned} \mathbf{f}_i = \mathsf{P}(Y = i) &= \sum_{j=1}^{60} \mathsf{P}(Y = i \mid T \in \mathcal{I}_j) \cdot \mathsf{P}(T \in \mathcal{I}_j) \\ &= \sum_{j=1}^{60} P_{ij}\, x_j. \end{aligned}$$

Note that $\boldsymbol{P}$ is a stochastic matrix with $\sum_{i=1}^{60} P_{ij} = 1$ for all j, reflecting the fact that any actual time in interval $\mathcal{I}_j$ must be rounded to some recorded value.

Coming to the empirical construction of the rounding matrix, the matrix formulation is introduced for two related reasons. First, it gives a compact representation of how actual one-minute intervals are mapped into recorded minutes. Second, the same matrix becomes the linear operator used later to reconstruct the full daily distribution. We denote by $\hat{\boldsymbol{P}}$ the matrix constructed from the estimated rounding weights $\hat{\boldsymbol{p}}$. We use sample counterparts, "replacing"the theoretical probability vector $\mathbf{f}$ with the empirical proportion vector $\boldsymbol{f} = (f_1, \ldots, f_{60})^\top$, where f_i is the observed proportion of events recorded at minute i. Moreover, under the uniformity assumption, we set $\boldsymbol{x} = (1/60, \ldots, 1/60)^\top$. The estimated matrix $\hat{\boldsymbol{P}}$ is constructed step by step from the estimated rounding weights. Let $\hat{\boldsymbol{p}} = (\hat{p}_1, \hat{p}_5, \hat{p}_{10}, \hat{p}_{15}, \hat{p}_{30}, \hat{p}_{60})^\top$ be the estimated vector of rounding weights. The one-minute component contributes to the diagonal of $\hat{\boldsymbol{P}}$:

$$\hat{P}_{ii} = \hat{p}_1 \qquad i = 1, \ldots, 60. \tag{9}$$

Then, for each coarser resolution $k > 1$ and for each attraction point $j \in R_k$, the estimated weight $\hat{p}_k$ is added to the row corresponding to the recorded value j and to the columns corresponding to actual-time intervals that would round to j at resolution k:

$$\hat{P}_{ji} = \hat{P}_{ji} + \hat{p}_k, \qquad i \in C_{jk}, \quad C_{jk} = \{i : y_k(i) = j\}. \tag{10}$$

The indices are interpreted circularly. For example, for $k = 10$ and attraction point $j = 30$, the contribution $\hat{p}_{10}$ is assigned to the actual-time intervals whose nearest 10-minute attraction point is 30. Boundary cases

halfway between two attraction points are split according to the same tie convention used in the rounding rule. This prevents double-counting and ensures that each column of $\hat{\boldsymbol{P}}$ sums to one, as required for a conditional probability matrix. This representation is consistent with the constrained least-squares decomposition above: if the matrix $\hat{\boldsymbol{P}}$ is constructed from the estimated rounding weights, the implied rounded distribution matches the fitted vector $\boldsymbol{\pi}(\hat{\boldsymbol{p}}_{\mathrm{LS}})$. The matrix gives a useful visual interpretation of the estimates and of their connection to clock time. In addition, we will use this matrix construction when we extend the analysis to a full day, in Section 3.

The matrix $\hat{\boldsymbol{P}}$ obtained starting from Boston crimes and some example conditional distributions are shown in Figure 2. The left panel illustrates how recorded (rounded) times relate to actual times. For a fixed actual time on the horizontal axis, the corresponding column shows all possible recorded values after rounding. Conversely, for a fixed recorded time on the vertical axis, the corresponding row shows the actual one-minute intervals that could have produced it. The illustrated matrix exhibits a block-banded structure: the diagonal band corresponds to minute-level recording, while progressively wider bands correspond to coarser rounding resolutions. The circular boundary is visible near minutes 60 and 1, where rounding to the hour can map times just before and just after the hour to the same recorded value. The right panel gives conditional distributions of actual times for selected recorded values. A recorded value such as 1 can only arise from a very narrow set of actual times, whereas recorded values such as 30 or 60 can be produced by several rounding resolutions and therefore correspond to much wider sets of possible actual times. This illustrates why the matrix formulation is useful: it makes explicit the pooling of actual times induced by each recorded rounded value.

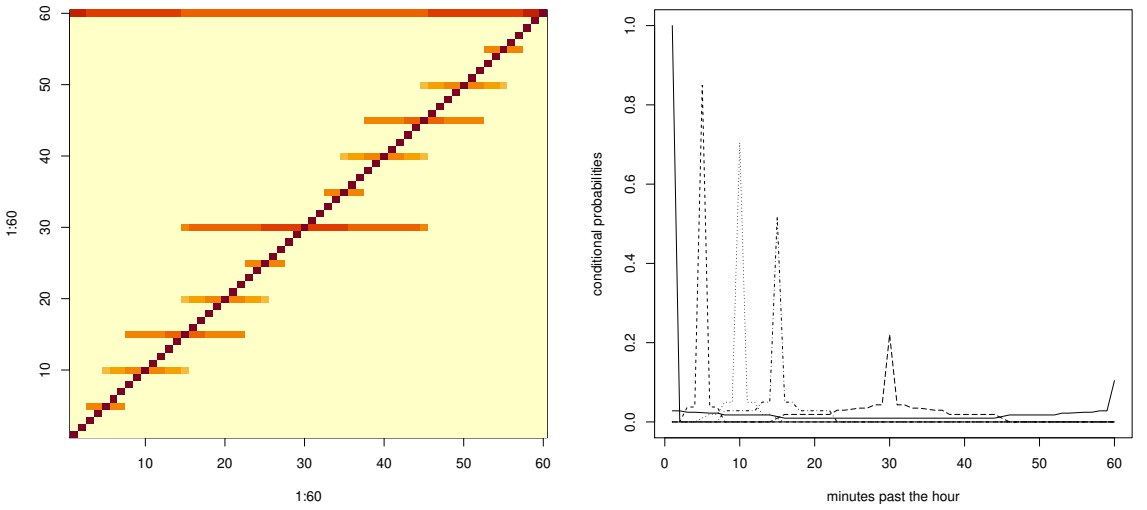

Fig. 2: Left: image representation of the estimated rounding matrix $\hat{\boldsymbol{P}}$ (raised to the power of 1/5 to aid visibility). The horizontal axis corresponds to the actual times, and the vertical axis to the recorded (rounded) times. Right: conditional distributions of actual values, given recorded values of $1, 5, 10, 15, 30$ and 60 (reading peaks from left to right).

2.5. Comparing rounding probabilities for different types of crime

The crime data are categorized by type of event, and so it may be of interest to see how the "rounding probability"changes with event type. Selecting those offence-code groups with frequency above 1000, we obtain the rounding probabilities from the minutes past the hour, as described above. The estimates of

$\hat{p}_k, k \in \mathcal{K}$ are shown in Table 2. All estimates are obtained under the simplex constraints. The asterisks identify the five offence groups for which the unconstrained least-squares solution would leave the simplex, with one estimated rounding weight negative. As a likelihood benchmark, constrained multinomial maximum likelihood gives identical estimates, to six decimal places, in all cases. We notice that this "crime"dataset

Offence Code Group	$\hat{p}_1$	$\hat{p}_5$	$\hat{p}_{10}$	$\hat{p}_{15}$	$\hat{p}_{30}$	$\hat{p}_{60}$
Aggravated Assault	0.820	0.052	0.017	0.037	0.047	0.027
Auto Theft	0.394	0.027	0.037	0.088	0.224	0.230
Auto Theft Recovery*	0.745	0.000	0.025	0.036	0.086	0.108
Commercial Burglary*	0.684	0.000	0.045	0.068	0.091	0.112
Confidence Games	0.471	0.032	0.013	0.026	0.104	0.352
Counterfeiting	0.686	0.015	0.019	0.037	0.050	0.193
Disorderly Conduct	0.799	0.054	0.025	0.057	0.034	0.031
Drug Violation	0.745	0.077	0.034	0.050	0.048	0.046
Fire Related Reports	0.958	0.007	0.014	0.003	0.007	0.012
Firearm Violations	0.895	0.046	0.007	0.011	0.012	0.030
Fraud	0.549	0.002	0.019	0.037	0.074	0.319
Harassment	0.649	0.034	0.006	0.047	0.095	0.168
Investigate Person	0.821	0.020	0.019	0.036	0.041	0.063
Investigate Property	0.847	0.014	0.011	0.017	0.053	0.057
Larceny	0.612	0.037	0.031	0.060	0.122	0.139
Larceny From Motor Vehicle	0.406	0.030	0.023	0.082	0.220	0.239
License Violation*	0.955	0.000	0.016	0.016	0.011	0.001
Liquor Violation*	0.907	0.053	0.000	0.029	0.004	0.009
Medical Assistance	0.924	0.019	0.009	0.015	0.014	0.020
Missing Person Located	0.625	0.014	0.028	0.058	0.143	0.132
Missing Person Reported	0.569	0.019	0.024	0.058	0.170	0.161
Motor Vehicle Accident Response	0.708	0.061	0.029	0.055	0.085	0.061
Other	0.778	0.041	0.020	0.042	0.060	0.059
Police Service Incidents*	0.800	0.000	0.003	0.054	0.076	0.068
Property Found	0.681	0.026	0.038	0.067	0.079	0.110
Property Lost	0.388	0.027	0.031	0.063	0.197	0.294
Recovered Stolen Property	0.726	0.038	0.030	0.047	0.066	0.092
Residential Burglary	0.568	0.032	0.027	0.079	0.153	0.141
Restraining Order Violations	0.820	0.021	0.012	0.022	0.054	0.072
Robbery	0.761	0.075	0.014	0.047	0.072	0.031
Simple Assault	0.795	0.037	0.028	0.048	0.050	0.040
Towed	0.913	0.000	0.018	0.022	0.025	0.022
Vandalism	0.574	0.023	0.021	0.049	0.151	0.181
Verbal Disputes	0.946	0.024	0.001	0.009	0.009	0.010
Violations	0.892	0.046	0.001	0.020	0.019	0.022
Warrant Arrests	0.809	0.047	0.026	0.054	0.027	0.037

Table 2. Estimated time-rounding probabilities for the most common offence groups. Estimates are obtained by constrained least squares as in Equation (6). Asterisks identify offence groups for which the non-negativity constraint is active, so that the solution lies on the boundary of the simplex.

contains events (e.g. Missing Person Located) which are not normally considered as crimes. As may be expected, the rounding probabilities depend very much on the type of event which was recorded. The estimates reveal three broad patterns:

- high-precision reporting ($\hat{p}_1 > 0.80$): events like *Fire Related Reports* ($\hat{p}_1 = 0.958$), *Verbal Disputes* ($\hat{p}_1 = 0.946$), *Firearm Violations* ($\hat{p}_1 = 0.895$) and *Aggravated Assault* ($\hat{p}_1 = 0.820$) show strong preference for minute-level recording. These incidents share common features: they typically occur in public view, involve immediate police or emergency response, and have multiple witnesses. Fire-related reports achieve near-perfect minute-level precision because they are dispatched through 911 calls with automatic time

stamping. Similarly, verbal disputes and firearm violations generate immediate calls for service, often from multiple witnesses, providing officers with precise timing information. The high $\hat{p}_1$ values indicate that reporting uncertainty is minimal, not because officers are more diligent, but because the incident circumstances provide genuine timing precision.

- High uncertainty incidents ($\hat{p}_1 < 0.45$): *Larceny From Motor Vehicle* ($\hat{p}_1 = 0.406$), *Auto Theft* ($\hat{p}_1 = 0.394$) and *Property Lost* ($\hat{p}_1 = 0.388$) exhibit substantial coarse rounding. Property crimes, particularly those involving vehicles or unattended belongings, exhibit the lowest timing precision, with 30-minute and 60-minute rounding probabilities often exceeding 20% each. This reflects genuine ignorance of the exact timing, consistent with criminological findings that crimes without witnesses or with delayed discovery often have poorly estimated times (see Klinger and Bridges (1997)). Clarke (2002) worked on theft from vehicles documenting that these are quintessential 'stealth offences' occurring when guardianship is absent and discovery is delayed. Our finding that $\hat{p}_{60}$ (hourly rounding) reaches 23-29% for these categories quantifies the timing uncertainty that criminologists have long recognized qualitatively.
- Extended-duration events ($\hat{p}_{60} > 0.30$): *Confidence Games* ($\hat{p}_{60} = 0.352$) and *Fraud* ($\hat{p}_{60} = 0.319$) show the highest hourly rounding rates. Unlike street crimes with discrete occurrence times, fraud and confidence schemes unfold over days or weeks. Victims often cannot identify a specific moment when the crime occurred—only when they discovered it. Recording such incidents at hour-level precision (or even day-level) is thus not measurement error but an accurate reflection that these crimes lack definable incident times.

We compare a selection of the crime events in Figure 3 in which the very different values of $\hat{p}_1$ are selected in the proportion of crimes recorded at 0 minutes past the hour. We show the corresponding density-scaled histograms for four common incident types in the Boston data. The observed patterns visually reflect the heterogeneity in rounding practices. *Larceny* shows pronounced peaks at multiples of 5, 15, 30, and 60 minutes, indicating coarse rounding and a correspondingly low $\hat{p}_1$. In contrast, the distribution for *Towed* is relatively flat, suggesting finer minute-level recording and a higher $\hat{p}_1$. *Drug Violation* and *Simple Assault* represent intermediate cases with moderate rounding peaks. The horizontal line in each panel marks the uniform density (1/60) expected for actual times of events in the absence of rounding bias. These results confirm that time rounding can occur for a variety of reasons. Sometimes it may be due to individual preference, whereas in the case of genuine uncertainty or ignorance, rounding to the nearest hour (for example) may be very common. It is clear, also, that Figure 1 contains a mixture of distributions, each of which may reflect a different set of rounding probabilities.

3. Extension to all times of the day

This section extends the proposed framework from within-hour rounding to the full daily cycle, allowing rounding behaviour to vary across hours of the day. We stress that the least-squares criterion appearing below has a different role from the first-stage rounding decomposition in Section 2.3. In the daily reconstruction it is used as a regularized inverse-problem criterion, combined with a smoothness penalty, rather than as an ordinary regression model or as a basis for regression-style inference. An initial approach could be to take the estimated within-hour matrix $\hat{\boldsymbol{P}}_h$ (where the new subscript recalls the dependence on the rounding of minutes past the hour) and extend this to the full day using the Kronecker product, giving the block diagonal (1440×1440) matrix $\hat{\boldsymbol{P}}_d = \boldsymbol{I}_{24} \otimes \hat{\boldsymbol{P}}_h$, where $\boldsymbol{I}_{24}$ is the 24×24 identity matrix. Then we could recover the distribution of actual times $\boldsymbol{x}_d = (x_1, \ldots, x_{1440})^\top$ by

$$\hat{\boldsymbol{x}}_d = \hat{\boldsymbol{P}}_d^{-1} \boldsymbol{f}_d \tag{11}$$

where $\boldsymbol{f}_d = (f_1, \ldots, f_{1440})^\top$ is the vector whose ith element denotes the proportion of recorded events occurring at i minutes past midnight ($i = 1, \ldots, 1440$). If the same within-hour matrix is used for all 24 hours, $\hat{\boldsymbol{P}}_d^{-1}$ is computed as $\boldsymbol{I}_{24} \otimes \hat{\boldsymbol{P}}_h^{-1}$; with hour-specific matrices, the inverse is instead block diagonal with blocks $(\hat{\boldsymbol{P}}^{(h)})^{-1}$, when these inverses exist. However, this approach to constructing $\hat{\boldsymbol{P}}_d$ is not correct, since, when considering a full day, those times rounded to the nearest hour would – in general – occur from 30 minutes before to 30 minutes afterwards, and not at any time in the hour before. This can be fixed by a

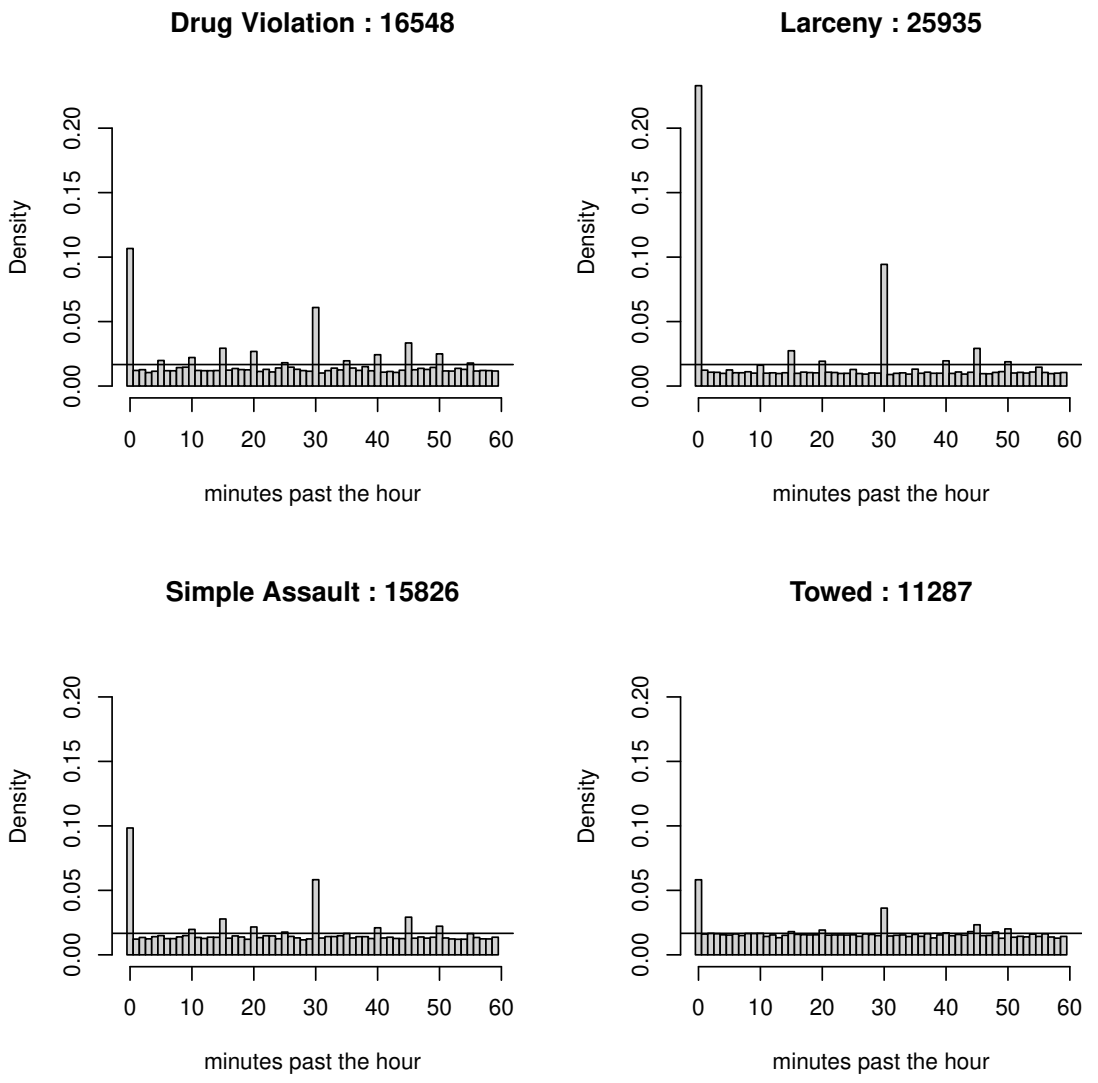

Fig. 3: Density histograms of recorded minutes past the hour for a selection of more common crime types. Sub-title reports the total number of crimes for each type. Horizontal lines indicate uniform distribution (1/60).

simple modification of $\hat{\boldsymbol{P}}_d$, including that the rounding probabilities for the half hours are shared to rounding to the hour before or afterwards — in a similar way to the discussion after Equation (10). Even after this correction, the naive unconstrained inverse problem remains ill-posed. Its solution has negative components, and a simple post-processing step that sets negative values to zero and renormalizes still produces a very noisy estimate. This motivates the penalized reconstruction used below.

We now tackle two of the issues which arose in the above approach. The most apparent problem is the roughness of the final solution, and we will later consider penalized approaches to smoothing. The second problem, which we now address, is that the assumption of time-invariant rounding probabilities may not

hold because temporal accuracy may vary during the day. Criminological studies suggest that the accuracy of incident timing depends on factors like witness availability and lighting conditions, which vary diurnally (see Klinger and Bridges (1997)). Crimes during night hours, when fewer people are awake to observe them, might therefore be reported with less accuracy. To overcome this issue, we estimate separate rounding parameters for each hour of the day. In principle, one could start with an adaptation of Equation (6) in which the left hand-side is replaced by $\boldsymbol{f}_d$ and the right-side is adapted to allow for an interaction between an hour-of-the-day factor and the rounding terms. Instead, we build $\hat{\boldsymbol{P}}_d$ as follows.

1. For each one hour block, let $\boldsymbol{f}^{(h)} = (f_{60h+1}, \ldots, f_{60h+60})^\top$ denote the extracted proportions for hour $h = 0, \ldots, 23$
2. Normalize $\boldsymbol{f}^{(h)}$ so that the entries sum to one (for $h = 0, \ldots, 23$)
3. Using the constrained criterion in Equation (6), obtain hour-specific rounding-weight estimates $\hat{p}_k^{(h)}$, $k \in \mathcal{K}$
4. Form the estimated hour-specific matrix $\hat{\boldsymbol{P}}^{(h)}$ from the hour-specific rounding weights $\hat{\boldsymbol{p}}^{(h)}$, using the same sequential procedure described in Section 2 using Equations (9) and (10)
5. Form a block-diagonal estimated matrix $\hat{\boldsymbol{P}}_d$ using $\hat{\boldsymbol{P}}^{(0)}, \ldots, \hat{\boldsymbol{P}}^{(23)}$
6. Adjust rows $60, 120, \ldots, 1440$ of $\hat{\boldsymbol{P}}_d$ as described above.

The above construction results in an estimated matrix $\hat{\boldsymbol{P}}_d$ in which the columns sum to one. However, we note an implicit assumption in the estimation of the $\hat{p}_k^{(h)}$ that the actual crime times are uniformly distributed within each hour. In periods when there is a more rapidly changing intensity of crimes this will only be approximately true. Then, when the inverse exists, we can recover the distribution of the actual times $\boldsymbol{x}_d = (x_1, \ldots, x_{1440})^\top$ by

$$\hat{\boldsymbol{x}}_d = \hat{\boldsymbol{P}}_d^{-1} \boldsymbol{f}_d \,.$$

The estimates $\hat{p}_k^{(h)}, k \in \mathcal{K}$, are shown in Figure 4, along with the estimated actual times $\hat{\boldsymbol{x}}_d$. We can note the large rounding probability to midnight which is shown in the final hourly value of $\hat{\boldsymbol{p}}$. The unsmoothed estimate (shown in the right panel of Figure 4) is quite noisy. But we note that allowing the rounding to vary by the hour of day has managed to smooth out the large peak at midnight (i.e. 1440 minutes past midnight) — whilst also accentuating the unusual spike at one minute past midnight. We note that the large spike at midnight is most likely due to "rounding to the day" — perhaps in the case that the time was completely unknown, and so a somewhat arbitrary midnight was selected. The large value at one minute past midnight (relative to one minute past any other o'clock) was perhaps used to eliminate any confusion about whether midnight should be allocated to the day before or the day after (since, for example, 00:00 on 4th March and 24:00 on 3rd March are the same time).

In real life, one would expect the density of actual times to be smooth. One could apply a "post-smoother"to $\hat{\boldsymbol{x}}_d$, for example, by employing a linear-circular local smoother, with a smoothing parameter chosen by cross-validation, but it would seem better if the smoothing could be done at the same time as the "inverse" operation. So, instead, to obtain a smooth solution, for given parameter λ, we minimize (over $\boldsymbol{x}_d$):

$$\|\boldsymbol{f}_d - \hat{\boldsymbol{P}}_d \boldsymbol{x}_d\|^2 + \lambda R(\boldsymbol{x}_d),$$

where $R(\boldsymbol{x}_d)$ denotes the "roughness"of $\boldsymbol{x}_d$. A natural choice for temporal smoothness is to penalize the squared second derivative, which on a circular time scale (midnight to midnight) corresponds to $R(\boldsymbol{x}_d) = \sum_{i=1}^{1440} \left(x_{i-1} - 2x_i + x_{i+1}\right)^2$, with indices taken modulo 1440. In matrix form, $R(\boldsymbol{x}_d) = \boldsymbol{x}_d^\top \boldsymbol{D}^\top \boldsymbol{D} \boldsymbol{x}_d$, where $\boldsymbol{D}$ is the 1440×1440 second-difference matrix with periodic boundary conditions, i.e

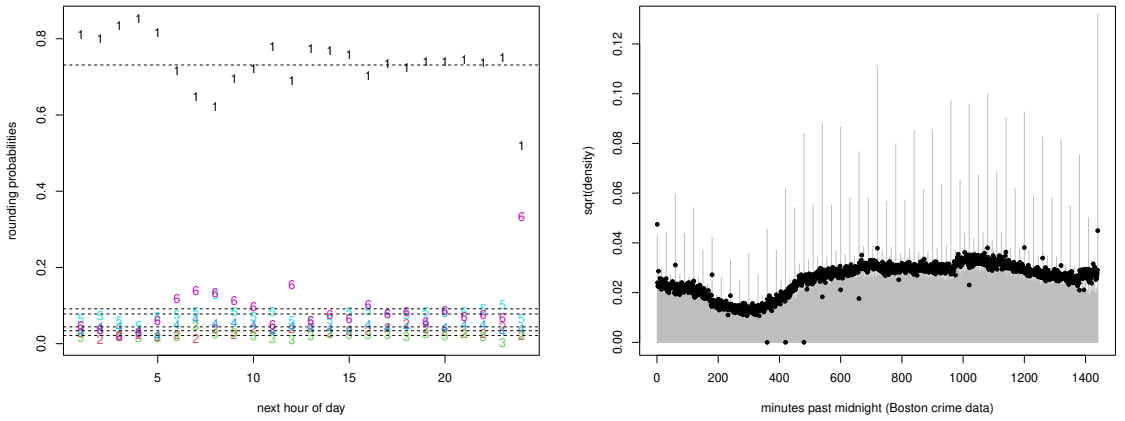

Fig. 4: Left: Rounding probabilities estimated by next hour of the day, where the symbols 1:6 correspond to $\hat{p}_1, \hat{p}_5, \hat{p}_{10}, \hat{p}_{15}, \hat{p}_{30}$ and $\hat{p}_{60}$ respectively. The horizontal lines correspond to the global estimates in Table 1. Right: Density scaled histogram of all recorded times (grey), naive solution given by the truncated and normalized version of $\hat{\boldsymbol{P}}_d^{-1}\boldsymbol{f}_d$ (black points), both shown on square root scale.

$$\boldsymbol{D}^\top\boldsymbol{D} = \begin{bmatrix} 6 & -4 & 1 & 0 & 0 & \cdots & 0 & 1 & -4 \\ -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 & 1 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & -4 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 6 & -4 & 1 \\ 1 & 0 & 0 & 0 & 0 & \cdots & -4 & 6 & -4 \\ -4 & 1 & 0 & 0 & 0 & \cdots & 1 & -4 & 6 \end{bmatrix}$$

Then, the penalized estimator is given by

$$\hat{\boldsymbol{x}}_d = \left(\hat{\boldsymbol{P}}_d^\top\hat{\boldsymbol{P}}_d + \lambda\boldsymbol{D}^\top\boldsymbol{D}\right)^{-1}\hat{\boldsymbol{P}}_d^\top\boldsymbol{f}_d. \tag{12}$$

When the smoothing penalty is introduced there is no longer any guarantee that $\sum x_i = 1$, but it is straightforward to normalize. An obvious way to select λ is by leave-one-out cross-validation (LCV). In ordinary regression, LCV removes the pair (x_i, f_i), refits the model, and then predicts the omitted response. Here the latent vector $\boldsymbol{x}_d$ is not observed; only the empirical rounded-time proportions $\boldsymbol{f}_d$ are observed. We therefore define LCV on the observed rounded scale as follows. Let $\boldsymbol{f}_d^{(-i)}$ denote the vector $\boldsymbol{f}_d$ with the ith component removed, and let $\hat{\boldsymbol{P}}_d^{(-i)}$ denote the matrix $\hat{\boldsymbol{P}}_d$ with the ith row removed. Let $\hat{\boldsymbol{x}}_d^{(-i)}$ be the penalized estimator obtained from Equation (12) using $\boldsymbol{f}_d^{(-i)}$ and $\hat{\boldsymbol{P}}_d^{(-i)}$. The fitted value for the omitted rounded-time component is then defined as

$$g_i = \left(\hat{\boldsymbol{P}}_d\hat{\boldsymbol{x}}_d^{(-i)}\right)_i,$$

where the full matrix $\hat{\boldsymbol{P}}_d$ is used only to evaluate the omitted component after fitting. After repeating this for all i, we obtain $\boldsymbol{g}_d = (g_1, \ldots, g_{1440})$. We truncate negative entries to zero, normalize $\boldsymbol{g}_d$ to sum to one, and select λ to minimize $\|\boldsymbol{g}_d - \boldsymbol{f}_d\|$.

We need only compute the above inverse in Equation (12) once, and then the inverse for the leave-one-out versions can be obtained by the Woodbury formula (Mardia et al., 2024) – however, the numerical minimization still takes a while, even with parallelization.

In Figure 5 we plot two smooth solutions, in both of which we use the values of λ chosen by cross-validation. One solution uses the rounding matrix $\hat{\boldsymbol{P}}_d$ in which the rounding probabilities within each hour are those given in Table 1. The other solution uses the $\hat{\boldsymbol{P}}_d$ as outlined in steps 1–6 above. The general pattern of crime rates is clear — there is a peak around 4:30pm and a minimum around 4:00am. The variable rounding matrix $\hat{\boldsymbol{P}}_d$ leads to a smoother solution, with a notably smaller estimate of the actual crime rate at midnight. This is due (in part) to the large estimate of $\hat{p}_{60}$ in the final hour, as seen in Figure 4 (Left panel). For the consistent rounding matrix $\hat{\boldsymbol{P}}_d$ there is a local peak around 2pm which may warrant further investigation.

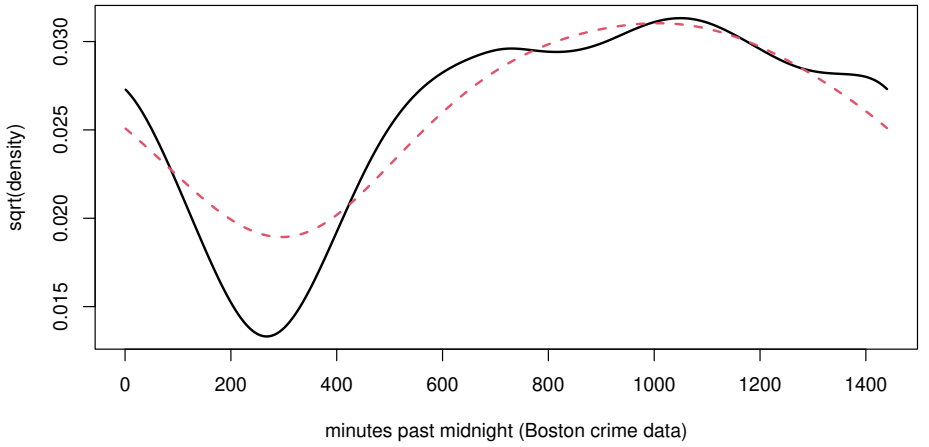

Fig. 5: Smooth solutions of the distribution of actual crime times during the day, with $\hat{\boldsymbol{x}}_d$ obtained by Equation (12), and λ chosen by cross-validation. Continuous line uses $\hat{\boldsymbol{P}}_d$ with consistent rounding probabilities over the hours of the day, and the dashed line for rounding probabilities estimated in each hour, both shown on square root scale.

4. Simulated data validation

We conduct a simple simulation experiment to illustrate the behaviour of the proposed estimator under controlled conditions. As an experiment, we will suppose that the density of actual events follows a sin pattern of the form $f(t) = 1 + \alpha \sin(2\pi t/1440)$ where $t \in [0, 1440)$ denotes the number of minutes past midnight. If we simulate a random sample from this distribution then, for each actual time, we generate a random recorded time according to the rounding probabilities in Table 1. We choose $\alpha = 0.1$ and record 99,692 times within the 24-hour period (we actually generated some actual times just before 0, and just after 1440 which were then rounded, but only used 'recorded' data in the range of one day). A histogram of the minutes past the hour, and the minutes of the day are shown in Figure 6. It is encouraging that the left panel resembles that of the real data in Figure 1. Applying the same constrained rounding-weight estimation procedure to the simulated recorded data gives the estimates reported in Table 3. These estimates are close to those obtained using the real data in Table 1 (which were used to simulate the data).

Rounding resolution k	Estimated rounding weight $\hat{p}_k$
1	0.74
5	0.03
10	0.02
15	0.04
30	0.07
60	0.10

Table 3. Estimated rounding weights for the simulated recorded data. Values are rounded to two decimal places.

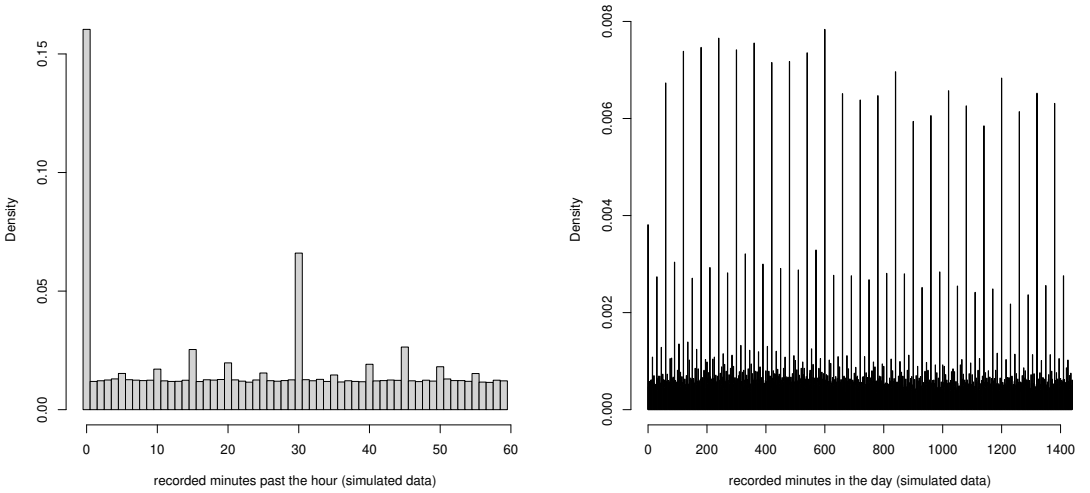

Fig. 6: Density-scaled histograms of simulated recorded data, when actual times follow a sin-distribution, and recorded times are rounded according to the estimated rounding weights $\hat{\boldsymbol{p}}$. Left: minutes past the hour; Right: minutes past midnight

Using the values $\boldsymbol{f}_d$ in the right panel of Figure 6, we estimate $\boldsymbol{x}_d$ using Equation (12), with λ chosen by cross-validation as described. For comparative purposes, we also tried directly smoothing the recorded times, using circular kernel estimation, with the smoothing parameter again chosen by cross-validation using the R library NPCirc (Oliveira et al., 2014).

When no smoothing is applied, but we simply use the rounding probabilities, estimated separately for each hour, then the solution is very rough (but non-negative for this well-behaved data), and the root mean squared error (RMSE) is 9.82×10^{-5}. When the smoothing parameter is selected by cross-validation, our proposed method of Section 3 achieves an RMSE of 1.45×10^{-5} when the rounding probabilities are consistent over the whole day, and 1.03×10^{-5} when rounding probabilities are estimated separately for each hour. From these values, as well as Figure 7, we can see much similarity in the two solutions. This is to be expected, as the data was simulated from a model with rounding probabilities which did not depend on the hour of the day. However, both of these solutions are much better when compared with 3.19×10^{-5} for the circular kernel estimator, corresponding to an error reduction by a factor of approximately 3.1. This substantial improvement indicates that explicitly correcting for rounding markedly reduces estimation bias under realistic, data-driven tuning. Figure 7 helps interpret the RMSE values reported above by revealing the bias structure of the competing estimators. The proposed penalized inversion approach closely tracks the true underlying density over the entire 24-hour cycle, whereas direct circular smoothing of the rounded

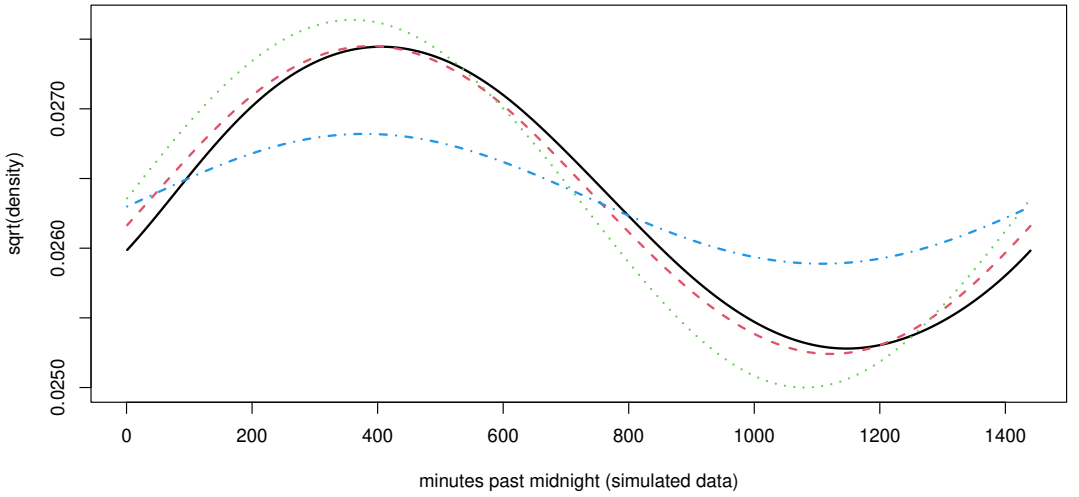

Fig. 7: Estimates of simulation model using rounded times, with smoothing parameters chosen by cross-validation: by means of Equation (12) in which (continuous) rounding probabilities are estimated separately for each hour of the day and (dashed) rounding probabilities are estimated for the whole day; (dot-dash) circular-linear kernel estimation based on the actual times; (dotted) the model used to simulate the data.

data exhibits systematic attenuation of the signal, particularly at peaks and troughs, due to unmodeled rounding effects. While deliberately simplified, these results confirm that explicitly accounting for rounding can substantially reduce bias relative to naive smoothing approaches.

5. Discussion and future work

The framework developed in this paper relies on a specific and interpretable class of rounding resolutions $\mathcal{K}$. While this choice is motivated by empirical evidence and parsimony, alternative rounding models could be considered. For instance, one could directly estimate separate weights for a larger collection of attraction points. Such models may reduce descriptive lack of fit, but they are less interpretable and less directly connected to common reporting resolutions. We therefore retain the parsimonious rounding-mixture specification for its clarity and direct insight into reporting behaviour. The analytical comparison with the multinomial likelihood in Section 2.3 further shows that the substantive estimates are not driven by the constrained least-squares criterion.

When applied to full daily-cycle data, additional rounding effects emerge beyond the minute scale. In particular, hours themselves appear to be rounded, as events for which the day is known but the time is uncertain are often recorded at salient values such as midnight. This phenomenon generates artificial spikes in the recorded data as depicted in Figure 4.

Despite these limitations, the estimated distributions consistently indicate that criminal activity is least frequent during the early morning hours, particularly between 4:00 and 5:00am. This pattern emerges clearly after correcting for rounding artifacts, underscoring the practical value of the proposed correction. The present paper is primarily descriptive and reconstructive: the first-stage constrained least-squares criterion is used to decompose the empirical minute-past-the-hour distribution into interpretable rounding components, and the multinomial likelihood is used as a likelihood benchmark under an independent-event sampling approximation. For the global minute-past-the-hour model, the two formulations coincide analytically under

the interior-solution condition verified in the data. A natural extension would be to develop formal uncertainty quantification, for example by using clustered resampling methods to account for possible dependence induced by common reporting practices, days, locations, operators, or offence types.

Acknowledgements

The authors would like to thank both referees and the Associate Editor for helpful and constructive comments which helped to improve the paper.

References

Bopp, M., and Faeh, D. (2008). End-digits preference for self-reported height depends on language. *BMC Public Health*, 8, 342.

Clarke, R. V. G. (2002). Thefts of and from cars in parking facilities. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Cohen, L. E., and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 588–608.

Di Marzio, M., Fensore, S., Panzera, A. Taylor, C.C. (2022). Density estimation for circular data observed with errors. *Biometrics*, 78, 248–260

Di Marzio, M., Fensore, S., and Taylor, C.C. (2023). Kernel regression for errors-in-variables problems in the circular domain. *Statistical Methods & Applications*, 32, 1217–1237

Farrington, D. P., and Welsh, B. C. (2002). Effects of improved street lighting on crime: a systematic review (Vol. 251). London: Home Office.

Fisher, N.I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press.

Giustinelli, P., Manski, C. F. and Molinari, F. (2022). Tail and center rounding of probabilistic expectations in the Health and Retirement Study. *Journal of Econometrics*, 231, 265–281.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. and Tourangeau, R. (2011). *Survey Methodology*. John Wiley & Sons.

Hand, D.J., Blunt, G., Kelly, M.G. and Adams, N.M. (2000). Data Mining for Fun and Profit. *Statistical Science*, 15, 111–131.

Hand, D.J., Blunt, G. (2001). Prospecting for gems in credit card data. *Journal of Management Mathematics*, 12 173–200.

Heitjan, D. F. and Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, 85, 304–314.

Jammalamadaka, S.R., SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.

Jammalamadaka, S. R., Ghosh, K. and Akiri, S. (2024). Testing uniformity on the circle using spacings when data are rounded. *Statistics & Probability Letters*, 210, 110114.

Klinger, D. A. and Bridges, G. S. (1997). Measurement error in calls-for-service as an indicator of crime. *Criminology*, 35, 705-726.

Ley, C., and Verdebout, T. (2017). *Modern Directional Statistics*. Chapman and Hall/CRC.

Ley, C., and Verdebout, T. (Eds.). (2018). *Applied Directional Statistics: Modern Methods and Case Studies*. CRC Press.

Mardia, K.V., Kent, J.T. and Taylor, C.C. (2024). *Multivariate Analysis (Second edition)*. Wiley.

Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2014). NPCirc: An R Package for Nonparametric Circular Methods. *Journal of Statistical Software*, 61, 1–26.

Sanko, N. and Iriguchi, N. (2022). Are self-reported times rounded? Insights from times reported by an objective third party. *Transportation Research Interdisciplinary Perspectives*, 16, 100698.

Sato, Y. and Maruyama, T. (2020). Modeling the rounding of departure times in travel surveys: comparing the effect of trip purposes and travel modes. *Transportation Research Record*, 2674, 628–637.

Sherman, L. W. and Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly*, 12, 625–648.

Stefanski, L.A., Carroll, L.J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 169–184.