MODELING TRANSPORT DATA UNDER STOCHASTIC AND LATENT CENSORSHIP

Inon Peled, Daniele Gammelli, Filipe Rodrigues, Dario Pacino, Francisco C. Pereira *

February 29, 2020

ABSTRACT

Data censorship involves two sets of corresponding values: known observations and latent (i.e., unknown) true values. Modeling of censored data has been researched in multiple works, including the famous Tobit model for known and deterministic censorship. However, when modeling demand for transport services, the censorship challenge becomes two-fold: not only is demand data inherently censored by limited supply, but it also typically lacks any account of the difference between observed and true demand. To address this problem, we devise and analyze two complementary methods for censored modeling, when no indication is given about the extent of censorship. The first modeling method is generic, while the other method is non-parametric and utilizes domain knowledge. Our experiments demonstrate both the importance of accounting for censorship and the capability of each method in reconstructing the underlying, latent patterns.

***K*eywords** censored modeling · stochastic censorship · latent censorship · transport demand

1 Introduction

Transport services require careful planning to satisfy highly volatile demand with limited supply. For example, transport service providers need to plan ahead how large the fleet shall be, where and when vehicles shall be deployed, and how much demand is expected to increase in the future. For optimal planning, reliable models of demand are thus needed to predict and meet user needs.

Transport demand modeling commonly relies on historical records of service usage to capture demand patterns. However, historical data about service usage is inherently limited by the historical supply that the service provider itself offered. Consequently, such data represents a biased, or *censored*, version of the underlying demand pattern that we wish to model.

If this inherent censorship is not accounted for, the resulting demand model will necessarily yield a biased estimate of demand and an inaccurate understanding of user needs, thereby leading to sub-optimal operational decisions. Furthermore, historical transport data commonly lacks explicit indication of which data records are censored and how intense the censorship is. Hence in the transport domain, censored modeling is both necessary and particularly challenging for accurate forecasting.

In this work, we construct and analyze two complementary approaches for modeling censored data, when no explicit information is given about the extent of censorship. The first modeling approach is generic, and we experiment it on synthetically generated data to examine how well it can reconstruct the underlying, true signal. The second approach is non-parametric and begins by estimating which observations are actually censored before modeling. We experiment with this alternative approach on real-world taxi demand data, as we further elaborate separately in (Gammelli et al. 2020)

* {inonpe, daga, rodr, darpa, camara}@dtu.dk. Danmarks Tekniske Universitet (DTU), Kgs. Lyngby, Denmark, 2800.

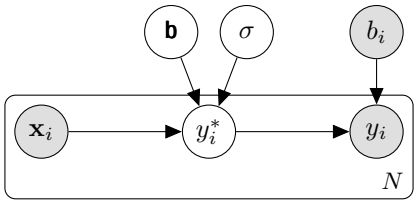


Figure 1: The Tobit model.

2 Related Work

Data censorship involves two sets of corresponding values: known observations and latent (i.e., unknown) true values. Each observation is either *non-censored*, so that it exactly equals the true value, or *censored*, so that it has been clipped at some threshold above or below the true value.

An early form of censored modeling is the Probit model (Aldrich et al. 1984) for 0/1 observations, which assumes that the probability of observing 1 depends linearly on the given explanatory features. James Tobin extended this to a popular model for censored regression (Tobin 1958), now called Tobit, where the censorship threshold is fixed and known. Tobit has also been extended through multiple variations (Greene 2011), such as: multiple latent variables (Amemiya 1985), count data (Terza 1985), Quantile Regression (J. L. Powell 1986), autoregressive Tobit (S. X. Wei 1999), and combination with Kalman Filter (Allik et al. 2015). Other methods of censored regression have also been suggested, predominantly for survival analysis, such as: Proportional Hazard models (Kay 1977), Accelerated Failure Time models (L.-J. Wei 1992); Regularized Linear Regression (Li et al. 2016); and Deep Neural Networks (Biganzoli et al. 2002; Wu et al. 2018).

A complementary approach to handle censored data is found in multiple data cleaning techniques. For instance, in the bike sharing domain, Jian et al. (2016) and Freund et al. (2019) focus on obtaining an unbiased estimate of arrival rate by omitting historical observations that are suspected as censored. In contrast, Albiński et al. (2018) replace censored observations with historical averages. Generally though, replacing and omitting censored observations can lead to loss of useful data, particularly when a large portion of observations are censored. It is thus more desirable to equip models with censorship awareness, thereby taking advantage of the whole data.

This work differs from the above works in several respects. First, we do not assume that the given dataset explicitly specifies which of the observations are censored, so that we treat censorship as stochastic and latent. Second, we offer a generic modeling scheme, which relies neither on data cleaning nor on a predefined functional form for the modeled phenomenon. In particular, we promote a non-parametric approach to modeling transport demand.

3 Methodology

3.1 Tobit Model for Deterministic and Known Censorship

The classic Tobit model concerns the following setting. We are given observations $y_1, \ldots, y_N$, which we assume to be independent. We also assume the existence of corresponding true values $y_1^*, \ldots, y_N^*$. For all $i = 1..N$, we say that y_i is *censored* if $y_i^* \neq y_i$, otherwise y_i is *non-censored* and $y_i^* = y_i$.

The true values y_i^* are *latent*, namely, they are neither given directly nor observed for censored y_i. We are, however, given also binary censorship labels $b_1, \ldots, b_n$, so that for all $i = 1..n$: $b_i = 0$ if y_i is non-censored, otherwise $b_i = 1$. For example, in a shared transport demand setting, y_i^* is the true, latent demand for shared mobility; the observed demand maintains $y_i \leq y_i^*$; and censorship is affected by the difference between actual demand and available supply.

Tobit parameterizes the dependency of y_i^* on explanatory features $\mathbf{x}_i$ through a linear relationship:

$$y_i^* \sim \mathcal{N}(\mathbf{b}^{\mathrm{T}}\mathbf{x}_i, \sigma^2)\,, \tag{1}$$

where $\mathbf{b}$ and σ are trainable parameters. This is illustrated in Fig. 1 as a Probabilistic Graphical Model, where nodes correspond to variables and parameters, edges are drawn from variables to conditionally dependent variables, and frames denote repetition (e.g., N times); shaded nodes correspond to observed variables, and all other nodes correspond to latent variables.

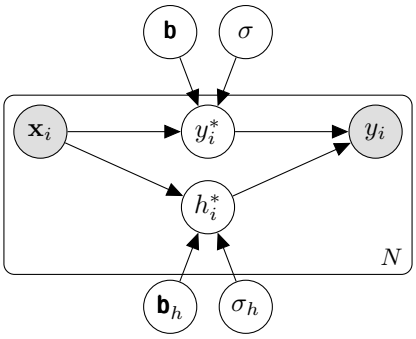


Figure 2: Model M^*.

To derive the likelihood function of Tobit, let us consider the case of deterministic upper censorship, where we are given some threshold h, and:

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* < h \\ h, & \text{if } y_i^* \geq h \end{cases} . \tag{2}$$

By Eq. (1) and Eq. (2):

1. If $b_i = 0$, then y_i is non-censored and so its likelihood is:

$$\frac{1}{\sigma}\varphi\left(\frac{y_i - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i}{\sigma}\right) , \tag{3}$$

   where φ is the standard Gaussian probability density function.

2. Otherwise, i.e., if $b_i = 1$, then y_i is censored and equal to h, hence its likelihood is:

$$1 - \quad \left(\frac{h - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i}{\sigma}\right) , \tag{4}$$

   where is the standard Gaussian cumulative density function.

Finally, because all observations are independent, their joint likelihood is:

$$\prod_{i=1}^{N} \left\{ \frac{1}{\sigma}\varphi\left(\frac{y_i - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i}{\sigma}\right) \right\}^{1-b_i} \left\{ 1 - \quad \left(\frac{h - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i}{\sigma}\right) \right\}^{b_i} , \tag{5}$$

which is a function of $\mathbf{b}$ and σ.

3.2 Model M^* for Completely Latent, Stochastic Censorship

Consider now a *completely latent censorship* setting, where we are given a censored dataset without any censorship labels or censorship thresholds. That is, we know that the given observations $y_1, \ldots, y_N$ are censored to some extent, but do not know which and how much. This setting is particularly prevalent when modeling mobility demand; for example, observations of public transport usage (e.g., taxi pickups and dropoffs, bus mounting and alighting, shared bike rentals and returns) are commonly given without explicit indication of how close they are to actual mobility demand.

Our goal is to model the completely censored data, e.g., to predict future observations. On one hand, we could ignore the censorship altogether, thus treating all observations as non-censored and equal to the latent, true values. However, such censorship-unaware modeling can severely underestimate the true, latent variable, as our experiments will show too. Therefore, we next develop and analyze censorship-aware models for better dealing with completely latent censorship.

First, we propose a generic model denoted M^*, as illustrated in Fig. 2. In M^*, variables $\mathbf{x}_i, y_i, y_i^*, \mathbf{b}_y, \sigma$ have the same roles as in Tobit; unlike Tobit, censorship thresholds $h_1^*, \ldots, h_N^*$ are latent random variables that depend on features $\mathbf{x}_i$ per parameters $\mathbf{b}_h, \sigma_h$, so that

$$h_i^* \sim \mathcal{N}(f(\mathbf{b}_h, \mathbf{x}_i), \sigma_h^2), \quad y_i^* \sim \mathcal{N}(g(\mathbf{b}, \mathbf{x}_i), \sigma^2), \quad y_i = \begin{cases} y_i^*, & y_i^* \leq h_i^* \\ h_i^*, & y_i^* > h_i^* \end{cases} , \tag{6}$$

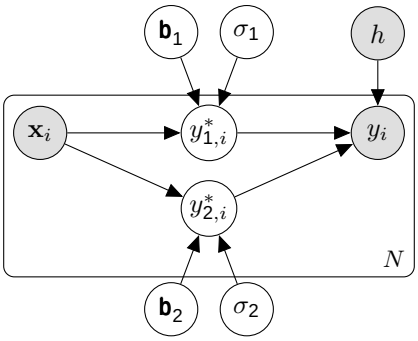


Figure 3: Tobit Type II model.

where f, g are freely chosen functional forms. For example, these forms can be linear as in Tobit, so that $f(\mathbf{b}_h, \mathbf{x}_i) = \mathbf{b}_h^{\mathrm{T}}\mathbf{x}_i$ and $g(\mathbf{b}, \mathbf{x}_i) = \mathbf{b}^{\mathrm{T}}\mathbf{x}_i$.

We now derive $\mathcal{L}^*$, the likelihood function of M^*. Let

$$z_i^* = y_i^* - h_i^* . \tag{7}$$

By Eq. (6), each observation y_i is censored if-and-only-if $z_i^* > 0$, namely,

$$y_i = \begin{cases} y_i^*, & z_i^* \leq 0 \\ h_i^*, & z_i^* > 0 \end{cases} . \tag{8}$$

Also, by properties of the Gaussian,

$$z_i^* \sim \mathcal{N}\left(f(\mathbf{b}_h, \mathbf{x}_i) - g(\mathbf{b}, \mathbf{x}_i), \sqrt{\sigma^2 + \sigma_h^2} \right) . \tag{9}$$

Hence by the law of total probability,

$$\begin{aligned} \Pr(y_i | \mathbf{b}, \mathbf{b}_h, \sigma, \sigma_h) &= \Pr(y_i | z_i^* > 0) \Pr(z_i^* > 0) + \Pr(y_i | z_i^* \leq 0) \Pr(z_i^* \leq 0) \\ &= \Pr(h_i^* = y_i) \Pr(z_i^* > 0) + \Pr(y_i^* = y_i) \Pr(z_i^* \leq 0) \\ &= \varphi_h(y_i)(1 - \quad_z(0)) + \varphi_y(y_i) \quad_z(0) , \end{aligned} \tag{10}$$

where φ_h is the PDF of h, φ_y is the PDF of y_i^*, and $\quad_z$ is the CDF of z_i^*. Finally, by independence of $y_1, \ldots, y_N$,

$$\begin{aligned} \mathcal{L}^*(y_1, \ldots, y_N, \mathbf{b}, \mathbf{b}_h, \sigma, \sigma_h) &= \prod_{i=1}^{N} \Pr(y_i | \mathbf{b}, \mathbf{b}_h, \sigma, \sigma_h) \\ &= \prod_{i=1}^{N} \left\{ \varphi_h(y_i)(1 - \quad_z(0)) + \varphi_y(y_i) \quad_z(0) \right\} . \end{aligned} \tag{11}$$

We conclude the description of M^* by comparing it to the commonly used Tobit Type II model, which is illustrated in Fig. 3. In Tobit Type II as well, observations depend on two latent random variables, so that

$$y_i = \begin{cases} y_{2,i}^*, & y_{1,i}^* \leq h \\ h, & y_{1,i}^* > h \end{cases} , \tag{12}$$

where h is a given threshold, as in the original Tobit model. $y_{1,i}^*$ thus controls whether y_i is censored or not, and in the latter case, $y_{2,i}^*$ separately controls the value of y_i. M^* thus differs from Tobit Type II in two respects: 1) in M^*, censorship thresholds are *stochastic* and *unknown*, 2) in M^*, both latent variables *symmetrically* control the observed values.

3.3 Gaussian Processes with Censored Likelihood

For non-parametric modeling of censored data, we now briefly describe Gaussian Processes (GPs) (Rasmussen and Williams 2005), which we also use later in Section 4.2. Given a dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with N input vectors $\mathbf{x}_i$ and

scalar outputs y_i, a Gaussian Process models the latent function that generates the data by defining a distribution over functions. This distribution is a multivariate Gaussian, defined by

$$p(\mathbf{f}^*|\mathbf{x_1}, \ldots, \mathbf{x_N}) = \mathcal{N}(\mathbf{m}, \mathbf{K}), \tag{13}$$

where $\mathbf{f}^* = [f_1^*, \ldots, f_N^*]^{\mathrm{T}}$ is a vector of latent random variables, $\mathbf{m}$ is a mean vector, and $\mathbf{K}$ is a covariance matrix with entries defined by a kernel function k, so that $\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$. The kernel $k(\mathbf{x}, \mathbf{x}')$ should properly reflect similarity between input vectors, hence for the data in our experiments in Section 4.2, we use a combination of the following two kernels:

1. Squared Exponential Kernel (SE):

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \lambda^2 \exp\left(-\frac{(\|\mathbf{x} - \mathbf{x}'\|)^2}{2\tau^2}\right), \tag{14}$$

   where $\|\cdot\|$ denotes Euclidean distance, and λ, τ are hyper-parameters.

2. Periodic Kernel:

$$k_{Per}(\mathbf{x}, \mathbf{x}') = \lambda^2 \exp\left(-\frac{2\sin^2(\pi \|\mathbf{x} - \mathbf{x}'\| / \rho)}{\tau^2}\right), \tag{15}$$

   where ρ is another hyper-parameter.

As is customary in GP modeling, we shall assume without loss of generality that the joint Gaussian distribution is centered on $\mathbf{m} \equiv \mathbf{0}$, and model each y_i as generated from a Gaussian distribution centered on f_i^* with noise variance σ^2. Also as common, we select kernel hyper-parameters in our experiments through Type-II Maximum Likelihood, also known as Empirical Bayes, whereby latent variables are integrated out. The likelihood function we use in GP experiments is

$$\prod_{i=1}^{N} \left\{ \frac{1}{\sigma} \varphi\left(\frac{y_i - f_i^*}{\sigma}\right) \right\}^{1-b_i} \left\{ 1 - \quad \left(\frac{y_i - f_i^*}{\sigma}\right) \right\}^{b_i}, \tag{16}$$

which we derive from Eq. (5) by replacing $\mathbf{b}^{\mathrm{T}}\mathbf{x}_i$, the Tobit prediction, with f_i^*, the GP prediction. We elaborate more on this likelihood function and its inference in (Gammelli et al. 2020), and so leave these details outside of this work.

4 Experiments

4.1 Experiments with M^*

As a first step in reasoning about the quality of modeling with M^*, we try out several artificially generated datasets, each featuring a different functional form of y^* and h^*. For each functional form, we also examine the effect of adding white noise ε_i independently to each of y_i^* and h_i^*. We experiment with $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon)$ for $\sigma_\varepsilon = 0$ (no noise), $\sigma_\varepsilon = 0.1$ (light noise), and $\sigma_\varepsilon = 0.3$ (medium noise). In each experiment, we fix $\sigma = \sigma_h = \sigma_\varepsilon$, so as to fit only $\mathbf{b}$ and $\mathbf{b}_h$. We then fit M^* using Maximum Likelihood Estimation (MLE), by minimizing the Negative Log-Likelihood (NLL) of $\mathcal{L}^*$, namely,

$$\mathrm{NLL}(\mathbf{b}) = -\sum_{i=1}^{N} \log\left\{ \varphi_h(y_i)(1 - \quad_z(0)) + \varphi_y(y_i) \quad_z(0) \right\}, \tag{17}$$

where $\mathbf{b}$ denotes all of $\mathbf{b}_y, \mathbf{b}_h$. Minimization is done via Powell optimizer (M. J. Powell 1964) with 100 random restarts.

The functional forms we experiment with are as follows.

1. *Linear*:

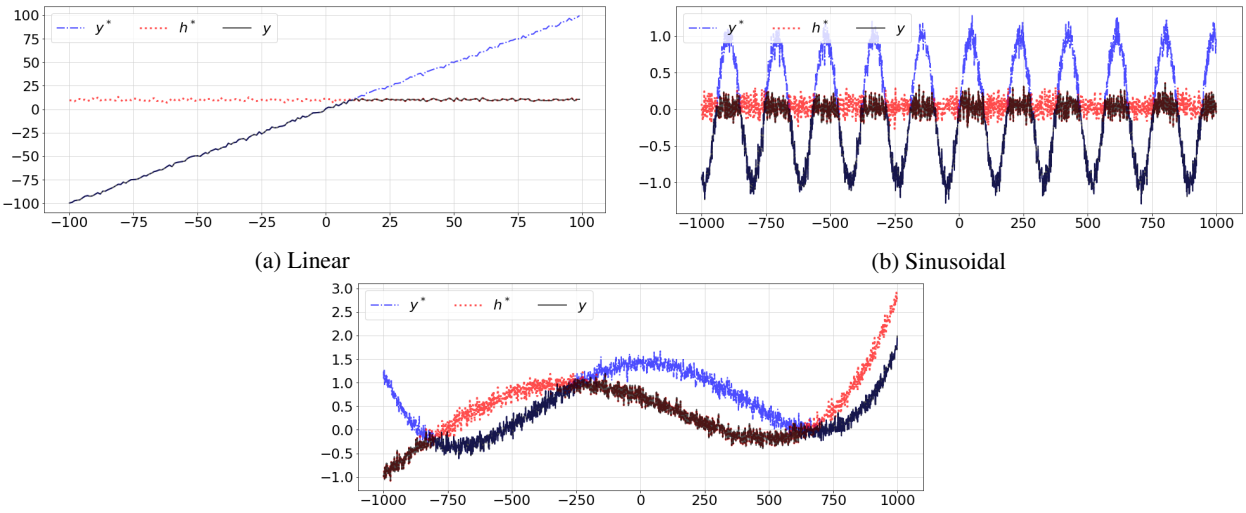
$$y^*(x) = \beta_{y,1}x + \beta_{y,0}, \quad h^*(x) = \beta_{h,1}x + \beta_{h,0}, \tag{18}$$

   where we choose $\beta_{y,1} = 1$, $\beta_{y,0} = 0$, $\beta_{h,1} = 0$, $\beta_{h,0} = 10$.

2. *Sinusoidal*:

$$y^*(x) = \sin\left(\frac{\beta_{0,y} + \beta_{1,y}x}{30}\right), \quad h^*(x) = \sin\left(\frac{\beta_{0,h} + \beta_{1,h}x}{30}\right), \tag{19}$$

   where each β is independently sampled from $\mathcal{N}(0, 3)$.



(a) Linear

(b) Sinusoidal

(c) Polynomial

Figure 4: Completely latent censorship for different functional forms with light noise.

3. *Polynomial*:

$$y^*(x) = \sum_{i=1}^{k} \beta_{i,y} T_i\left(\frac{x}{1000}\right), \quad h^*(x) = \sum_{i=1}^{k} \beta_{i,h} T_i\left(\frac{x}{1000}\right), \tag{20}$$

where k is a given polynomial degree, $T_i(x)$ denotes the i'th Chebyshev polynomial, and each β is independently sampled from $\mathcal{N}(0, 1)$. We experiment with $k = 5$.

We illustrate these forms in Fig. 4 for light noise. In all experiments, $\mathbf{x} = [-1000, -999, \ldots, 999]$.

We evaluate the performance of M^* through two measures, both pertaining to the difference between the parameters $\hat{\mathbf{b}}$ of fitted M^* and the actual parameters $\mathbf{b}^*$ which the dataset uses. The first measure is NLL as in Eq. (17), and the second measure is Maximum Absolute Error (MxAE), defined as:

$$\text{MxAE} = \ell_\infty\left(\left|\hat{\mathbf{b}} - \mathbf{b}^*\right|\right). \tag{21}$$

The experimental results are given in Table 1. Under zero noise, $\hat{\mathbf{b}}$ of M^* is nearly identical to the actual $\mathbf{b}^*$. As noise gradually increases, so does $\hat{\mathbf{b}}$ become increasingly different from $\mathbf{b}^*$ Also, as the functional form becomes more complex and the noise increases, NLL($\hat{\mathbf{b}}$) increasingly drops below NLL($\mathbf{b}^*$). This suggests that as the underlying latent signals become more complex and noisy, the original parameters may no longer account for the highest possible likelihood of the observed data.

Form	**Noise**	**NLL($\hat{\mathbf{b}}$)**	**NLL($\mathbf{b}^*$)**	**MxAE**
Linear	None	-737.246	-737.246	2.21e-7
	Light	270.264	270.458	0.042
	Medium	734.494	735.866	3.808
Sinusoidal	None	-7367.317	-7367.317	3.27e-4
	Light	-1785.655	-1779.856	0.328
	Medium	404.872	411.252	1.104
Polynomial	None	-7369.691	-7369.744	3.98e-4
	Light	-1808.441	-1785.326	0.059
	Medium	304.161	392.614	1.918

Table 1: M^* Performance

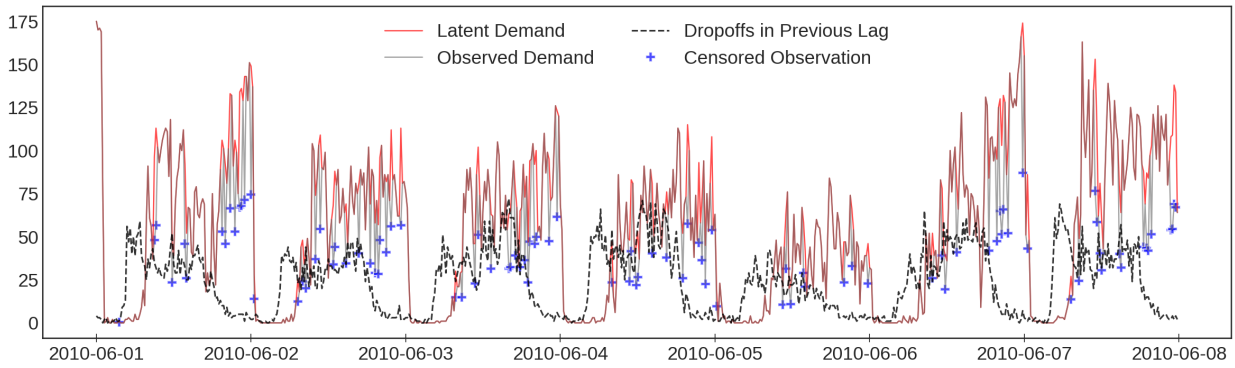


Figure 5: Example of selected censorship labels for $\gamma = 0.1, c = 0.5$.

4.2 Experiments with Gaussian Processes

Model M^* is designed to handle a general setting of latent censorship, where no censorship labels are assumed. Let us now present another approach for handling this setting by taking advantage of domain knowledge to estimate censorship labels before fitting models. To demonstrate this alternative approach, we use a case study of short-term transport demand, where we estimate censorship labels per available supply.

The data for this case study is pickups and dropoffs of yellow hail-based taxis (Donovan and Work 2014) within an approximately $1\,\text{km} \times 1\,\text{km}$ squared area near LaGuardia airport in New York City. The true pickup counts, $\mathbf{y}^* = [y_1^*, \ldots, y_{672}^*]$, are aggregated per $15\,\text{min}$ consecutive intervals in the first week of June 2010. At that time, NYC Taxis were almost entirely of the yellow type and thus accounted for virtually all ride-hailing demand.

Using this dataset, we wish to estimate the effectiveness of our approach under varying extent of censorship. For this, we employ the following scheme, which both labels observations as censored and, for the sake of experimentation, decreases the values of censored observations. For every time step $i = 1..672$, the scheme uses d_{i-1}, the number of dropoffs observed in the previous lag (i.e., taxi supply), to stochastically censor y_i^*, the true pickups (i.e., taxi demand), as follows.

First, each censorship label $b_i \in \{0, 1\}$ is treated as a Bernoulli variable with success probability

$$\left\{1 + \exp\left(\ln\left(\frac{1-\gamma}{\gamma}\right) - \frac{y_i - d_{i-1}}{y_i}\right)\right\}^{-1}, \tag{22}$$

so that approximately $0 < \gamma < 1$ of all y_i^* are labeled as censored. Next, each observation for which $b_i = 1$ is set to y_i to $(1 - c)y_i^*$, where $0 \leq c \leq 1$ represents unsatisfied ride-hailing demand. Fig. 5 illustrates an example of this censorship scheme for a certain combination of γ, c.

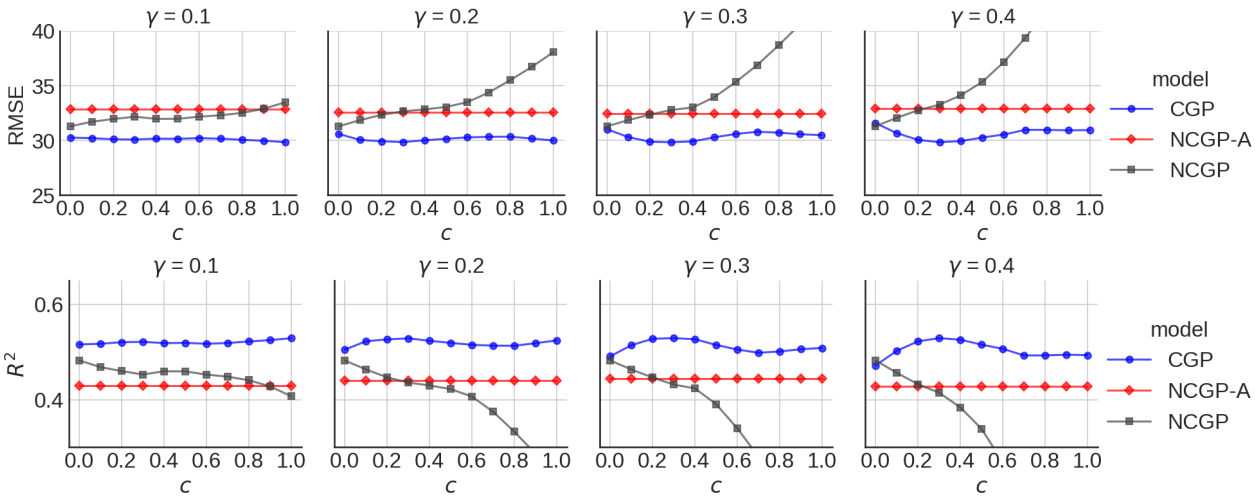
We experiment with $\gamma = 0.1, 0.2, 0.3, 0.4$ and $c = 0, 0.1, \ldots, 1$. For these values of γ, the average percentages of censored observations are, respectively, $13\%, 23\%, 32\%, 40\%$ – close to the expected values. The censorship magnitude increases with c, so that $c = 0$ and $c = 1$ represent limiting cases: when c tends to 0, censored observations become very close to true values, whereas when c tends to 1, censored observations are zeroed.

In these experiments, we compare between three Gaussian Process models:

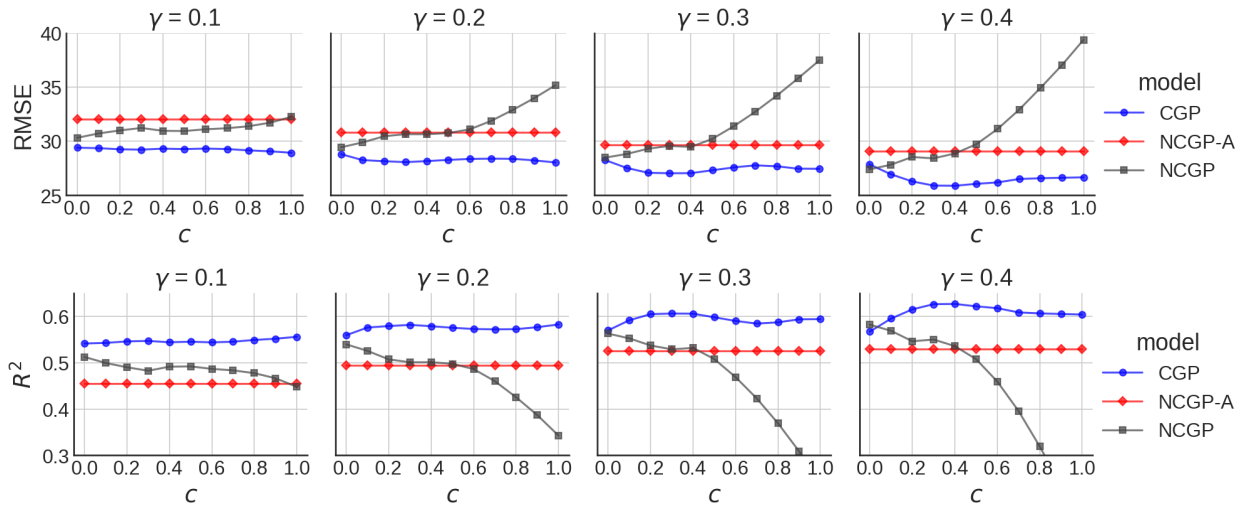
1. *Non-Censored Gaussian Process (NCGP)*, which completely ignores the existence of censorship, as is quite common in literature.
2. *Non-Censored Gaussian Process, Aware of censorship (NCGP-A)*: similar in function to NCGP, but trained only on non-censored points. NCGP-A thus avoids exposure to a censored, biased version of the true demand, but does so at the price of potentially losing useful information.
3. *Censored Gaussian Process (CGP)*: this model considers all observations – censored and non-censored – through the likelihood function defined in Eq. (16).

For all models, the explanatory features are $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{672}]$, where for all time steps $i = 1 \ldots 672$, $\mathbf{x}_i$ consists of: i, the corresponding hour-of-day in $0 \ldots 23$, and the corresponding day-of-week in $0 \ldots 6$. For NCGP-A and CGP, $\mathbf{x}_i$ also contains the binary censorship label b_i. The GP kernel for all three models is

$$k(\mathbf{x}, \mathbf{x}') = k_{SE}(\mathbf{x}, \mathbf{x}') + k_{Per}(\mathbf{x}, \mathbf{x}'), \tag{23}$$



(a) Evaluation over all observations.



(b) Evaluation over only non-censored observations.

Figure 6: Performance of Gaussian Process models with censored likelihood. Extreme values of NCGP are omitted for easier comparison between NCGP-A and CGP.

per Eq. (14) and Eq. (15).

All models are trained via BFGS optimizer (Fletcher 2013) for at most 1000 iterations, starting from the same initial kernel hyper-parameters. Because our censorship scheme is stochastic, we experiment with each γ, c combination independently a total of 30 times. In each independent experiment, we fit and evaluate through cross-validation with 21 time-consecutive folds, each consisting of 32 observations over 8 hours. For each γ and c, we evaluate by comparing the latent pickup counts to the predicted means over all 30 experiments, both on the entire dataset and exclusively on non-censored observations.

The results are illustrated in Fig. 6. As γ and c increase, so do the percentage of censored observations and the intensity of censorship. Hence by ignoring censorship, NCGP deteriorates rapidly as the censored observations draw it downwards, away from the latent pickups. The performance of NCGP-A does not depend on censorship intensity, because NCGP-A discards of all censored observations before fitting. NCGP-A maintains a rather stable performance as γ increases, and so apparently has enough non-censored points for constructing a stable fit.

Finally, CGP takes advantage of the censorship labels, so that it often reconstructs the latent signal better than the censorship-unaware models do. In particular, CGP is the best performing model when evaluated on only non-censored

observations, which implies that CGP is also the more reliable model when observations are known to accurately reflect the latent ground truth.

In conclusion, we note that we have also experimented with a similarly sized spatial area in the middle of Manhattan, NYC. Contrary to the time series used in this Section, the Manhattan cell exhibited a repetitive and regular demand pattern, for which NCGP-A and CGP performed quite closely. The noticeably better performance of CGP in this Section thus suggests that the advantages of censored modeling emerge in more challenging settings — where the ability to extract meaningful information from censored observations is indeed essential for capturing the underlying demand pattern.

5 Summary and Future Work

Demand modeling is a fundamental building block in numerous decision making processes, and commonly relies on extrapolating knowledge from historical data. A reliable demand prediction model must take into consideration censoring, particularly so when demand is implicitly limited by supply. Censored modeling is especially challenging in the transport domain, where datasets often lack explicit information about which records are censored and how intense the censorship is.

We have devised two complementary approaches for dealing with this challenge of latent censorship. The first approach is generic, and our experiments on synthetic data show that it can reconstruct the latent signals for various functional forms with light to medium noise. The second approach uses domain knowledge to reconstruct censored labels before fitting models, and we have demonstrated its effectiveness on real-world transport data through non-parametric modeling with Gaussian Processes. Our experiments also show that models which ignore data censorship are prone to yield a biased estimation of the underlying, latent transport demand.

Both of our approaches thus leverage the information embedded in censored data, rather than discard it through cleaning techniques. For future work, we plan to apply the generic modeling approach also to real-world datasets of transport demand. We further plan to expand the non-parametric modeling approach to multiple spatial areas and utilize their spatio-temporal correlations.

Acknowledgement

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Individual Fellowship H2020-MSCA-IF-2016, ID number 745673.

References

Albiński, S., P. Fontaine, and S. Minner (2018). "Performance analysis of a hybrid bike sharing system: A service-level-based approach under censored demand observations". In: *Transportation Research Part E: Logistics and Transportation Review* 116, pp. 59–69. ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2018.05.011.

Aldrich, J. H., F. D. Nelson, and E. S. Adler (1984). "The Linear Probability Model". In: *Linear probability, logit, and probit models*. Vol. 45. Sage. Chap. 1, pp. 9–27.

Allik, B., C. Miller, M. J. Piovoso, and R. Zurakowski (2015). "The Tobit Kalman filter: an estimator for censored measurements". In: *IEEE Transactions on Control Systems Technology* 24.1, pp. 365–371.

Amemiya, T. (1985). "Tobit Models". In: *Advanced Econometrics*. Harvard university press, p. 387.

Biganzoli, E., P. Boracchi, and E. Marubini (2002). "A general framework for neural network models on censored survival data". In: *Neural Networks* 15.2, pp. 209–218.

Donovan, B. and D. Work (Dec. 2014). *New York City Taxi & Limousine Data*. https://uofi.app.box.com/v/NYCtaxidata/folder/2332218797. Accessed: 2019-Apr-29.

Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.

Freund, D., S. G. Henderson, and D. B. Shmoys (2019). "Bike Sharing". In: *Sharing Economy: Making Supply Meet Demand*. Ed. by M. Hu. Cham: Springer International Publishing, pp. 435–459. ISBN: 978-3-030-01863-4. DOI: 10.1007/978-3-030-01863-4_18.

Gammelli, D., I. Peled, F. Rodrigues, D. Pacino, H. A. Kurtaran, and F. C. Pereira (2020). "Estimating Latent Demand of Shared Mobility through Censored Gaussian Processes". In: *arXiv preprint arXiv:2001.07402*.

Greene, W. H. (2011). "Censored Data and Truncated Distributions". In: *SSRN Electronic Journal*, pp. 695–734. DOI: 10.2139/ssrn.825845.

Jian, N., D. Freund, H. M. Wiberg, and S. G. Henderson (Dec. 2016). "Simulation optimization for a large-scale bike-sharing system". In: *2016 Winter Simulation Conference (WSC)*, pp. 602–613. DOI: 10.1109/WSC.2016.7822125.

Kay, R. (1977). "Proportional hazard regression models and the analysis of censored survival data". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 26.3, pp. 227–237.

Li, Y., B. Vinzamuri, and C. K. Reddy (2016). "Regularized weighted linear regression for high-dimensional censored data". In: *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, pp. 45–53.

Powell, J. L. (1986). "Censored regression quantiles". In: *Journal of econometrics* 32.1, pp. 143–155.

Powell, M. J. (1964). "An efficient method for finding the minimum of a function of several variables without calculating derivatives". In: *The computer journal* 7.2, pp. 155–162.

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press. ISBN: 026218253X.

Terza, J. V. (1985). "A Tobit-type estimator for the censored Poisson regression model". In: *Economics Letters* 18.4, pp. 361–365.

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables". In: *Econometrica* 26.1, pp. 24–36. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1907382.

Wei, L.-J. (1992). "The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis". In: *Statistics in medicine* 11.14-15, pp. 1871–1879.

Wei, S. X. (1999). "A Bayesian approach to dynamic Tobit models". In: *Econometric Reviews* 18.4, pp. 417–439.

Wu, W., M.-Y. Yeh, and M.-S. Chen (2018). "Deep censored learning of the winning price in the real time bidding". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 2526–2535.