

A survival analysis approach to understanding travel time decay

Aaron Weinstock

Renaissance Planning
316 Holland St. | Durham, NC 27701
a.weinstock@citiesthatwork.com

November 27, 2019

Contents

1	Introduction	2
2	Data	2
3	Modeling methods	3
3.1	The Cox proportional hazards model	3
3.2	Selection of covariates	4
3.3	Initial models	5
3.4	The proportional hazards assumption	5
3.4.1	Time-dependent models	6
3.5	Other diagnostics	6
3.5.1	Outliers	6
3.5.2	Linearity of continuous covariates	7
4	Model fitting	7
4.1	Personal vehicle	7
4.2	Non-motorized	14
4.3	Transit	18
5	Operationalizing the model	18
5.1	Deriving a decay factor	18
5.1.1	Trip duration component	19
5.1.2	Change in conditions component	19
5.1.3	Calculating decay	20
5.2	A walkthrough example	21
6	Conclusion	23

1 Introduction

Historically, travel time decay curves have been estimated via exponential regression on time. First, for each time t in a set of travel times $t = 1, 2, \dots, t_{max}$, number of trips lasting at least t minutes would be calculated. Then, assuming the functional form $y = Ae^{bt}$, a regression would be fit with t (time) as the sole explanatory variable and y (the number of trips lasting at least t minutes) as the response.

While this method does elucidate the underlying decay curve, it is a naive approach in that it assumes travel time decay relies only on time. By considering time as the only covariate, this type of model ignores the impact other variables could have on travel time decay. For example, increased age could be associated with shorter tripmaking for walking trips, and thus a steeper travel time decay. Thus, a more informative travel time decay model may seek to include the effects of both time and other relevant covariates. This would not only produce a more realistic travel time decay curve, but allow for a quantification of the relationship between included covariates and travel time decay.

This paper suggests the use of a survival analysis approach to modeling travel time decay. Survival analysis encompasses a set of statistical methods interested in modeling the time to an “event” using a set of predictor variables. It is most often seen in clinical trials, where a researcher may be interested in how long patients survive after being given a particular treatment. It can be extended to travel time decay by considering the ending of a trip as the “event” of interest. Table 1 provides the travel time decay analogs of a more common survival analysis example.

Table 1: Travel time decay in a survival analysis context

	Common survival analysis	Analogs in trip decay
Observational unit	Disease patients at a hospital	Trip records
Event	Death	Ending of a trip
Predictor variables	Treatment, age, sex	Financial cost, purpose
Meaning of "survival"	How long the patient survives	How long the trip lasts

2 Data

The data comes from the Massachusetts Travel Survey (MTS), conducted by the Massachusetts Department of Transportation and published in June 2012. It was provided by the Metropolitan Area Planning Council (MAPC), which serves Boston, MA and its metropolitan region.

The data includes 190,215 trip records from 37,023 persons across 15,033 households in Massachusetts. The trip dataset includes attributes such as duration, mode, purpose, cost, destination, etc. The person dataset includes attributes such as age, education, employment status, etc. The household dataset includes size, number of workers, number of vehicles, income, location, etc.

This analysis required known values for all model variables, so some trip records were filtered out prior to model fitting based on survey responses. In addition, models were fit individually on different classes of modes, so different filtering conditions were necessary to create each model set. Table 2 details the modes used for each model, the conditions on which model variables were filtered, and the resulting number of trip records used for model fitting.

Table 2: Trip record filtering protocol for each mode-specific model

Model	Modes	Variable	Filter condition	Total records
Personal vehicle	Driver or passenger of auto, truck, van, or motorcycle	Trip purpose	Not "other" or "while traveling – other"; not a "loop trip" with no secondary purpose	
		Parking cost	Not "do not know" or "refused"	
		Toll cost	Not "do not know" or "refused"	112,262
		Household size	Not "do not know" or "refused"	
Non-motorized	Walk or bike	Trip purpose	Not "other" or "while traveling – other"; not a "loop trip" with no secondary purpose	
		Household size	Not "do not know" or "refused"	26,137

3 Modeling methods

3.1 The Cox proportional hazards model

Of particular interest for travel time decay modeling is the Cox proportional hazards model (henceforth referred to as the Cox model). The Cox model is, in essence, a multivariate regression. It attempts to model the *hazard function* $h(t)$ using a set of predictor variables. The hazard function can then be used to calculate the *survival function* $S(t)$, or the probability that a trip lasts longer than t . Though hazard is *not* a probability, it is helpful to think of hazard as the probability that a trip will end in an infinitesimally small time window $[t, t + \delta t]$, provided that the trip has lasted until time t . In other words, it is the instantaneous risk of a trip ending, given that it has lasted to that point. The mathematical relationship between the survival function and the hazard function is $S(t) = e^{-\int_0^t h(y)dy}$, where $\int_0^t h(y)dy$ is the *cumulative hazard function*, or the total accumulated risk up to time t . Details are left out for clarity, but this is based on the definition of $h(t)$ as a limit of event probability as $\delta \rightarrow 0$.

The Cox model is multiplicative, in that it assumes a multiplicative relationship between the covariates and the hazard. It takes the following form:

$$h(t|x_{i1}, \dots, x_{ip}) = h_0(t) * e^{\beta_1 x_{i1} + \dots + \beta_p x_{ip}}$$

Where:

- t is time,

- $h(t|x_i)$ is the hazard function for trip i with covariates x_{i1}, \dots, x_{ip} ,
- $h_0(t)$ is the baseline hazard function,
- β_j is the effect of the j^{th} covariate, and
- x_{ij} is the value of the j^{th} covariate for the i^{th} trip.

An important feature of the Cox model is that *the baseline hazard function is not specified*. This is due to the *proportional hazards assumption*: the ratio of hazards between two unique observations should **not** rely on time. For example, if after 15 minutes, a walking trip taken by a 60 year old carries twice the risk of ending as a walking trip taken by a 20 year old, it should also carry twice the risk after 10 minutes, 30 minutes, or indeed any other time. Non-proportional hazards may be thought of as an interaction between a covariate and time. Because hazards are all proportional under the proportional hazards assumption, the baseline hazard is arbitrary.

The Cox model is “semi-parametric”, in that it requires no assumptions about the baseline hazard function. Parametric methods can be used if the baseline hazard itself is of interest, and these methods indeed rely on specifying a distribution for the baseline hazard (common selections include the Exponential, Weibull, and Lognormal distributions). Though these models have more power when this distribution is correctly specified, they are not robust to misspecifications, and thus are highly risky. For this particular application of travel time decay curves, the baseline hazard was not of particular interest, and it sufficed to assume a general decrease in trip survival with time in the absence of covariate information (as is the outcome of Cox modeling). However, it stands to reason that understanding the baseline hazard may be of interest in other travel time decay modeling applications.

3.2 Selection of covariates

Covariates to be considered for each mode-specific model were selected from the set of available attributes in the trip, person, and household datasets reviewed in the Data section. Importantly, only covariates that would be (1) known for a given trip/person/household in a predictive setting or (2) easy to estimate for a given trip/person/household were selected for the model. These conditions were established for easier use of final model in practice. For example, though it is easy to see how age may be a relevant predictor of travel time decay for walking trips (older people may tend to take shorter trips, for example), it is unlikely a trip taker’s age is a readily available data point in a predictive setting. Furthermore, attempts to predict age for inclusion in a final model could be complicated and add uncertainty to travel time decay estimates.

Table 3 details the candidate covariates considered for each mode-specific model.

Table 3: Candidate covariates considered for each mode-specific model

Model	Covariates
Personal vehicle	Trip purpose
	Parking cost
	Toll cost
	Household size
Non-motorized	Trip purpose
	Household size

In all models, trip purpose was collapsed from 24 categories provided in the MTS data to three broad “purpose groups”: home, work, and non-work. Table 4 details how purposes were collapsed. “Loop trips”

were classified according to their secondary purpose (which were also coded according to Table 4), or filtered out if no secondary purpose was provided.

Table 4: Methodology for collapsing purpose to home, work, and non-work groups

Model	MAPC survey code	MAPC purpose definition
Home	1	Working at home
	2	All other home activities
Work	3	Work/job
	4	All other activities at work
Non-work	12	Work business related
	5	Volunteer work/activities
	6	Attending class
	7	All other school activities
	8	Changed type of transportation
	9	Drop off passenger from car
	10	Pick up passenger from car
	13	Service private vehicle
	14	Routine shopping
	15	Shopping for major purchases
	16	Household errands
	17	Personal business
	18	Eat meal outside of home
	19	Health care
	20	Civic/religious activities
	21	Outdoor recreation/entertainment
	22	Indoor recreation/entertainment
	23	Visit friends/relatives

3.3 Initial models

Initial model fitting began by including all candidate covariates for each mode-specific model a linear form (see Table 3). Backward selection was applied to create a “statistically valid” model – one with only significant covariates. Backward selection involves iteratively eliminating the least significant covariate from the model [from the set of insignificant covariates], and refitting the model until all remaining covariates are significant. If all variables were significant in the first model fit, backwards selection was not applied, as there was no need for it.

3.4 The proportional hazards assumption

Though there are multiple assumptions underlying the Cox model, by far the most important assumption is that of proportional hazards (see Section 3.1, The Cox proportional hazards model). Without this assumption, the Cox model is uninterpretable and completely invalid. Exploration of model diagnostics thus

began with the proportional hazards assumption, and proceeded to other assumptions only once proportional hazards was verified.

The proportional hazards assumption was tested for each covariate using the *Schoenfeld Residuals Test* (also known as the *Schoenfeld Individual Test*). The null hypothesis of this test is that time and scaled Schoenfeld residuals are uncorrelated, which should be true in the case of proportional hazards. The alternative hypothesis is that time and scaled Schoenfeld residuals are correlated. Thus, a significant result indicates that a lack of proportional hazards - this would indicate the model results are dependent on time. If no time dependence was observed in any covariate at $\alpha = 0.05$, diagnostics proceeded to validation of other assumptions. Otherwise, a new model accounting for time dependence was fit.

3.4.1 Time-dependent models

Multiple strategies exist to account for time dependence in Cox models. One such method is the use of “time-dependent coefficients”, which involves unique effect sizes for covariates depending on the period of time. Essentially, time-dependent coefficients create a step function of survival time, where each covariate is time-independent within each step. This is particularly relevant and interpretable in the context of travel time decay, because it allows for a sense of “bracketing” trip times, and understanding unique effects across travel times. Again, consider the theoretical use of age as a covariate in predicting walking travel time decay. For very short trips, there is likely a minimal effect of age (if any), as all people are likely equally comfortable walking less than five minutes. For longer trips, however, the effect size might increase, as young adults may be comfortable walking 20-30 minutes while senior citizens are not.

To fit a time-dependent coefficients model, variables for which the Schoenfeld Residuals test was significant at $\alpha = 0.05$ were explored visually using plots of the scaled Schoenfeld residuals against time. This graphical exploration demonstrated the functional form of the effect of these covariates over time, informing decisions on potential break points for intervals in which covariates may be time-independent. Interval break points were also subject to the researcher’s discretion of an “interpretable interval”: in other words, intervals would be created on 5 or 10 minute break points rather than odd-valued break points which would have less practical significance.

Once interval break points were identified, the data was reconfigured with dummy entries to reflect the selected intervals. For example, if trip T lasted 16 minutes, and intervals of 0-10 minutes and 10-20 minutes were identified from the plots, trip T would be split into two entries T_1 and T_2 . Trip T_1 would start at 0 minutes, go until 10 minutes, and would not end (i.e. would not experience an “event”); trip T_2 would start at 10 minutes, go until 20 minutes, and end somewhere in that interval (i.e. would experience an “event”). A time-dependent model was then fit on this reconfigured “interval” data, and the Schoenfeld Residuals Test was again applied to observe if any time dependence remained. This process was repeated until all time dependence was eliminated, yielding a model which adhered to the proportional hazards assumption.

3.5 Other diagnostics

In addition to the proportional hazards assumption, the Cox model also makes assumptions similar to those of standard regression. For this analysis, diagnostics for the assumptions of (1) no outliers in the data and (2) proper functional form of continuous covariates (which, in this case, was linear) were explored. These two assumptions were checked only after proportional hazards had been verified, either in a standard model or in a time-dependent coefficients model.

3.5.1 Outliers

Outliers were explored using a plot of the deviance residuals against the index of the residuals. Unusually large positive or negative deviance residuals were indicative of an outlier (large positive-valued deviance

residuals indicate a trip that ended far earlier than expected according to the model, while large negative-valued residuals indicate a trip that ended far later than expected according to the model). Trips that were considered outliers by this visual exploration were looked at individually for an inclusion/exclusion decision.

Given that the trips all came from the same survey and the same state, outlier points would rarely be considered erroneous or apart from the population of interest, and thus would only be excluded in a very extreme case.

3.5.2 Linearity of continuous covariates

Linearity of continuous covariates was explored using plots of martingale residuals of a “null Cox model” (including only one covariate) against that covariate. If the relationship between that covariate and hazard is truly linear, then a linear trend should be evident in the plot; otherwise, a different functional form may represent a better fit. If necessary, better fits were explored by fitting a new “null Cox model” with a different functional form of the covariate.

Like with outliers, the functional form of a variable would be changed only in a very extreme case, where the deviation from a linear relationship was severe or the use of a nonlinear relationship was obvious. The use of linearity between a covariate and hazard rate does not necessarily make the model *wrong* if it is not the true relationship, but rather indicates that a better fit may be available. However, estimating the true functional form could be a time-consuming challenge if the relationship is especially complicated, and such an endeavor is likely unnecessary for this particular analysis. In addition, linear relationships will be more easily interpretable, which is a primary goal of this model.

4 Model fitting

The model fitting results are detailed individually for each mode-specific model.

4.1 Personal vehicle

The model including all candidate covariates produced a statistically valid fit, so no backward selection was completed. Parking cost, toll cost, household size, and purpose were all significant to hazard rate (the purpose “home” was used as the reference level for fitting the purpose covariate). The model coefficients are provided in Table 5, while the hazard ratios are provided in Table 6.

Table 5: Coefficients for initial personal vehicle model

Covariate	β	$SE(\beta)$	z	$Pr(> z)$
Parking cost	-0.0346	0.0038	-9.0058	< 0.0001
Toll cost	-0.1189	0.0106	-11.2057	< 0.0001
Household size	0.0490	0.0028	17.2631	< 0.0001
Purpose: work	-0.3109	0.0123	-25.2405	< 0.0001
Purpose: non-work	0.2127	0.0085	24.8810	< 0.0001
Origin TAZ parking	-0.0270	0.0012	-22.2987	< 0.0001
Destination TAZ Parking	-0.0151	0.0014	-10.8966	< 0.0001
Origin TAZ terminal time	-0.0148	0.0047	-3.1707	0.0015
Destination TAZ terminal time	-0.0373	0.0040	-9.3298	< 0.0001

The coefficients should be interpreted as follows:

- For a continuous covariate X , assuming all other covariates are held constant, $\beta_X < 0$ indicates that hazard rate – or the risk of the trip ending – decreases as X increases, which means longer travel times and slower travel time decay for larger values of X . Conversely, $\beta_X > 0$ indicates that hazard rate increases as X increases, which means shorter travel times and faster travel time decay for larger values of X .
- For a categorical covariate Y with levels a, ref (where ref is the reference level), assuming all other covariates are held constant, $\beta_{Y_a} < 0$ indicates that hazard rate is higher in group a than in the reference group, which means longer travel times and slower travel time decay in a relative to ref . Conversely, $\beta_{Y_a} > 0$ indicates that hazard rate is lower in group a than in the reference group, which means shorter travel times and faster travel time decay in a relative to ref .

Table 6: Hazard ratios (HR) for initial personal vehicle model

Covariate	HR	Inverse HR	95% CI for HR
Parking cost	0.9660	1.0352	[0.9587, 0.9733]
Toll cost	0.8879	1.1262	[0.8697, 0.9066]
Household size	1.0503	0.9522	[1.0444, 1.0561]
Purpose: work	0.7328	1.3647	[0.7153, 0.7507]
Purpose: non-work	1.2370	0.8084	[1.2165, 1.2579]
Origin TAZ parking	0.9733	1.0274	[0.971, 0.9756]
Destination TAZ Parking	0.9850	1.0152	[0.9823, 0.9877]
Origin TAZ terminal time	0.9853	1.0149	[0.9763, 0.9944]
Destination TAZ terminal time	0.9634	1.0380	[0.9558, 0.9709]

The hazard ratios should be interpreted as follows:

- For a continuous covariate X , assuming all other covariates are held constant, the increase of X to $X + 1$ results in a change in the hazard rate by a factor of HR_X .
- For a categorical covariate Y with levels a, ref (where ref is the reference level), assuming all other covariates are held constant, group a has a hazard rate HR_{Y_a} times that of group ref .

Schoenfeld Residuals Tests on the covariates revealed time dependence in all covariates. The results of the test are shown in Table 7 (the **Global** row contains the results of a global test on the model, which tests if there is any time dependence in the model as a whole). A time-dependent coefficients model was explored as a result of present time dependence.

Table 7: Results of Schoenfeld Residuals Tests for initial personal vehicle model

Covariate	ρ	χ^2	p
Parking cost	0.0207	71.4861	< 0.0001
Toll cost	0.0286	311.1182	< 0.0001
Household size	-0.0293	56.1660	< 0.0001
Purpose: work	0.0572	216.8715	< 0.0001
Purpose: non-work	-0.0304	61.8114	< 0.0001
Origin TAZ parking	0.058	235.9203	< 0.0001
Destination TAZ Parking	0.0281	53.6765	< 0.0001
Origin TAZ terminal time	0.0232	35.1154	< 0.0001
Destination TAZ terminal time	0.0222	32.8805	< 0.0001
Global		1645.7382	< 0.0001

Exploration of scaled Schoenfeld residuals vs. time plots followed by iterative model fitting and re-checking of the proportional hazards assumption resulted in a time-dependent coefficients model with break points at 5, 10, 20, and 30 minutes. This model featured no time dependence, and thus this model was used going forward. The model coefficients are provided in Table 8, while the hazard ratios are provided in Table 9. Interpretation of the coefficients and the hazard ratios stay the same, with the exception that the interpretation is only valid in the specified time interval.

Table 8: Coefficients for time-dependent personal vehicle model

Covariate	Interval	β	SE(β)	z	Pr(> z)
Parking cost	0-5 minutes	-0.1433	0.0276	-5.1836	< 0.0001
Parking cost	5-10 minutes	-0.1387	0.0215	-6.4651	< 0.0001
Parking cost	10-20 minutes	-0.0700	0.0104	-6.7235	< 0.0001
Parking cost	20-30 minutes	-0.0434	0.0089	-4.8799	< 0.0001
Parking cost	> 30 minutes	-0.0089	0.0042	-2.0947	0.0362
Toll cost	0-5 minutes	-2.5828	0.3690	-6.9998	< 0.0001
Toll cost	5-10 minutes	-2.7670	0.3100	-8.9262	< 0.0001
Toll cost	10-20 minutes	-0.4139	0.0376	-11.0170	< 0.0001
Toll cost	20-30 minutes	-0.0602	0.0158	-3.8103	1e-04
Toll cost	> 30 minutes	-0.0206	0.0074	-2.7956	0.0052
Household size	0-5 minutes	0.0627	0.0063	9.9612	< 0.0001
Household size	5-10 minutes	0.0666	0.0057	11.7325	< 0.0001
Household size	10-20 minutes	0.0494	0.0053	9.3447	< 0.0001
Household size	20-30 minutes	0.0027	0.0080	0.3354	0.7373
Household size	> 30 minutes	0.0103	0.0080	1.2823	0.1997
Purpose: work	0-5 minutes	-0.3973	0.0342	-11.6116	< 0.0001
Purpose: work	5-10 minutes	-0.5854	0.0297	-19.7332	< 0.0001
Purpose: work	10-20 minutes	-0.4509	0.0237	-19.0651	< 0.0001
Purpose: work	20-30 minutes	-0.1710	0.0294	-5.8086	< 0.0001
Purpose: work	> 30 minutes	0.0452	0.0271	1.6678	0.0953
Purpose: non-work	0-5 minutes	0.3224	0.0190	16.9749	< 0.0001
Purpose: non-work	5-10 minutes	0.1945	0.0168	11.6049	< 0.0001
Purpose: non-work	10-20 minutes	0.2153	0.0158	13.6302	< 0.0001
Purpose: non-work	20-30 minutes	0.1652	0.0244	6.7781	< 0.0001
Purpose: non-work	> 30 minutes	0.0569	0.0251	2.2657	0.0235
Origin TAZ parking	0-5 minutes	-0.0711	0.0068	-10.5313	< 0.0001
Origin TAZ parking	5-10 minutes	-0.0665	0.0049	-13.5377	< 0.0001
Origin TAZ parking	10-20 minutes	-0.0346	0.0024	-14.2290	< 0.0001
Origin TAZ parking	20-30 minutes	-0.0158	0.0024	-6.5516	< 0.0001
Origin TAZ parking	> 30 minutes	-0.0054	0.0020	-2.7531	0.0059
Destination TAZ parking	0-5 minutes	-0.0411	0.0064	-6.4410	< 0.0001
Destination TAZ parking	5-10 minutes	-0.0355	0.0045	-7.8201	< 0.0001
Destination TAZ parking	10-20 minutes	-0.0228	0.0029	-7.9561	< 0.0001
Destination TAZ parking	20-30 minutes	-0.0132	0.0030	-4.3588	< 0.0001
Destination TAZ parking	> 30 minutes	-0.0040	0.0024	-1.6169	0.1059
Origin TAZ terminal time	0-5 minutes	-0.0487	0.0120	-4.0745	< 0.0001
Origin TAZ terminal time	5-10 minutes	-0.0395	0.0103	-3.8458	1e-04
Origin TAZ terminal time	10-20 minutes	0.0154	0.0085	1.8174	0.0691
Origin TAZ terminal time	20-30 minutes	0.0203	0.0117	1.7347	0.0828
Origin TAZ terminal time	> 30 minutes	0.0264	0.0113	2.3333	0.0196
Destination TAZ terminal time	0-5 minutes	-0.0776	0.0104	-7.4730	< 0.0001
Destination TAZ terminal time	5-10 minutes	-0.0387	0.0087	-4.4326	< 0.0001
Destination TAZ terminal time	10-20 minutes	-0.0337	0.0075	-4.4959	< 0.0001
Destination TAZ terminal time	20-30 minutes	-0.0097	0.0101	-0.9621	0.336
Destination TAZ terminal time	> 30 minutes	-0.0020	0.0092	-0.2194	0.8263

Table 9: Hazard ratios (HR) for time-dependent personal vehicle model

Covariate	Interval	HR	Inverse HR	95% CI for HR
Parking cost	0-5 minutes	0.8665	1.1540	[0.8208, 0.9148]
Parking cost	5-10 minutes	0.8705	1.1488	[0.8346, 0.9079]
Parking cost	10-20 minutes	0.9324	1.0725	[0.9135, 0.9516]
Parking cost	20-30 minutes	0.9575	1.0444	[0.9409, 0.9743]
Parking cost	> 30 minutes	0.9912	1.0089	[0.983, 0.9994]
Toll cost	0-5 minutes	0.0756	13.2344	[0.0367, 0.1557]
Toll cost	5-10 minutes	0.0628	15.9116	[0.0342, 0.1154]
Toll cost	10-20 minutes	0.6611	1.5127	[0.6142, 0.7116]
Toll cost	20-30 minutes	0.9416	1.0620	[0.9129, 0.9712]
Toll cost	> 30 minutes	0.9796	1.0208	[0.9656, 0.9939]
Household size	0-5 minutes	1.0647	0.9392	[1.0517, 1.0779]
Household size	5-10 minutes	1.0689	0.9356	[1.057, 1.0808]
Household size	10-20 minutes	1.0506	0.9518	[1.0398, 1.0615]
Household size	20-30 minutes	1.0027	0.9973	[0.9871, 1.0185]
Household size	> 30 minutes	1.0103	0.9898	[0.9946, 1.0264]
Purpose: work	0-5 minutes	0.6722	1.4877	[0.6286, 0.7188]
Purpose: work	5-10 minutes	0.5569	1.7958	[0.5254, 0.5902]
Purpose: work	10-20 minutes	0.6370	1.5698	[0.6082, 0.6673]
Purpose: work	20-30 minutes	0.8428	1.1865	[0.7956, 0.8929]
Purpose: work	> 30 minutes	1.0462	0.9558	[0.9921, 1.1032]
Purpose: non-work	0-5 minutes	1.3804	0.7244	[1.33, 1.4328]
Purpose: non-work	5-10 minutes	1.2147	0.8232	[1.1755, 1.2553]
Purpose: non-work	10-20 minutes	1.2402	0.8063	[1.2024, 1.2792]
Purpose: non-work	20-30 minutes	1.1796	0.8477	[1.1246, 1.2374]
Purpose: non-work	> 30 minutes	1.0585	0.9447	[1.0077, 1.112]
Origin TAZ parking	0-5 minutes	0.9314	1.0737	[0.9191, 0.9438]
Origin TAZ parking	5-10 minutes	0.9357	1.0688	[0.9267, 0.9447]
Origin TAZ parking	10-20 minutes	0.9660	1.0352	[0.9614, 0.9706]
Origin TAZ parking	20-30 minutes	0.9843	1.0160	[0.9796, 0.989]
Origin TAZ parking	> 30 minutes	0.9946	1.0054	[0.9908, 0.9984]
Destination TAZ parking	0-5 minutes	0.9597	1.0420	[0.9478, 0.9718]
Destination TAZ parking	5-10 minutes	0.9651	1.0361	[0.9566, 0.9737]
Destination TAZ parking	10-20 minutes	0.9775	1.0230	[0.972, 0.983]
Destination TAZ parking	20-30 minutes	0.9869	1.0133	[0.981, 0.9928]
Destination TAZ parking	> 30 minutes	0.9961	1.0040	[0.9913, 1.0008]
Origin TAZ terminal time	0-5 minutes	0.9524	1.0499	[0.9304, 0.975]
Origin TAZ terminal time	5-10 minutes	0.9613	1.0403	[0.9422, 0.9808]
Origin TAZ terminal time	10-20 minutes	1.0155	0.9847	[0.9988, 1.0325]
Origin TAZ terminal time	20-30 minutes	1.0205	0.9799	[0.9974, 1.0442]
Origin TAZ terminal time	> 30 minutes	1.0268	0.9739	[1.0042, 1.0498]
Destination TAZ terminal time	0-5 minutes	0.9253	1.0807	[0.9067, 0.9443]
Destination TAZ terminal time	5-10 minutes	0.9621	1.0394	[0.9457, 0.9787]
Destination TAZ terminal time	10-20 minutes	0.9668	1.0343	[0.9527, 0.9812]
Destination TAZ terminal time	20-30 minutes	0.9904	1.0097	[0.971, 1.0101]
Destination TAZ terminal time	> 30 minutes	0.9980	1.0020	[0.9801, 1.0162]

The plot of deviance residuals for the updated, time-dependent model is shown in Figure 1. Though a few potential outliers are observed in the “trip ended earlier than expected” category (large positive deviance residuals), they are small in number, and individual exploration revealed no grounds on which to exclude any points as outliers.

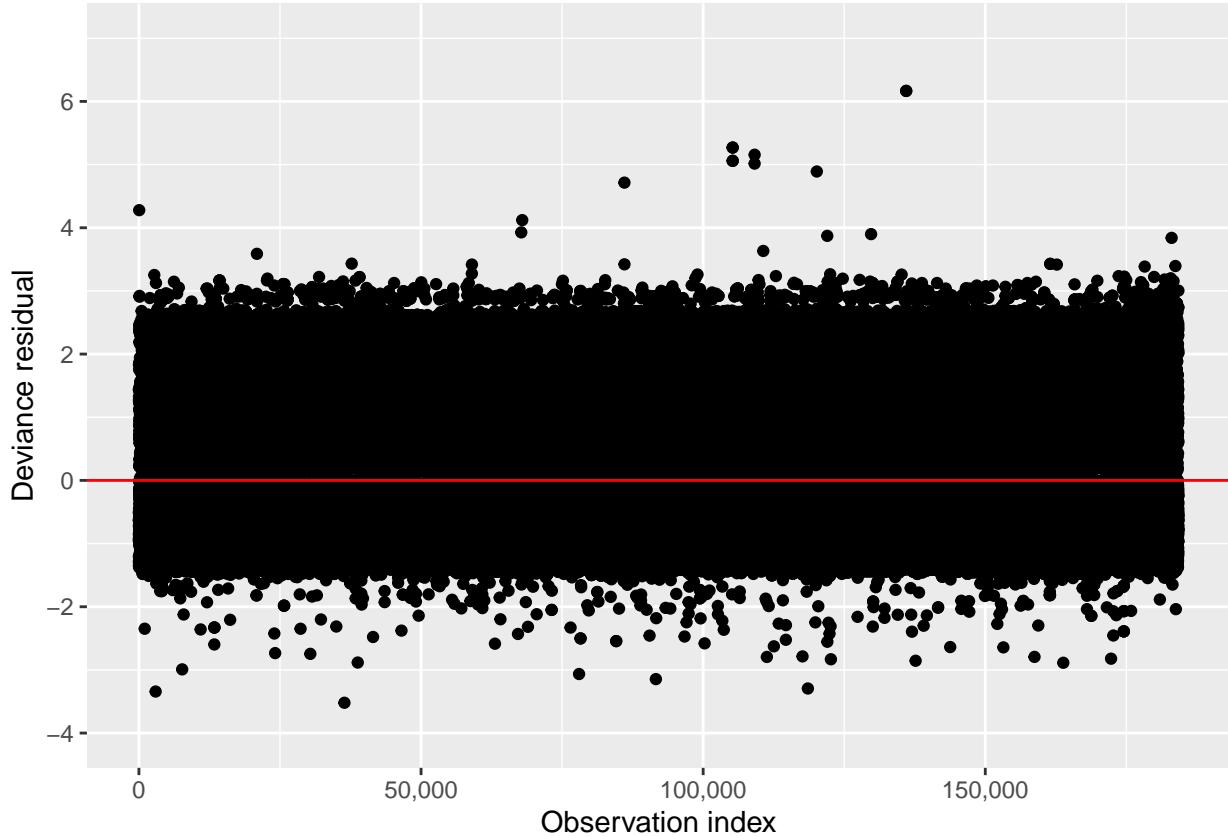


Figure 1: Time-dependent personal vehicle model deviance residuals

Plots of null Cox model Martingale residuals against parking cost, toll cost, household size, origin and destination TAZ parking, and origin and destination TAZ terminal time revealed no drastic deviations from linearity, so no functional forms were changed in the time-dependent model. However, exact linearity did not appear to be achieved for any of these covariates, indicating better functional forms likely exist for the inclusion of these covariates in the model. For this analysis, linearity was deemed “close enough” in the context of time and processing limitations.

Having estimated and validated a Cox proportional hazards model for personal vehicle travel time decay, the estimated survival curve was compared to the sample curve for the “most common set of conditions”. For personal vehicle travel, the most common set of conditions was:

- Parking cost = \$0
- Toll cost = \$0
- Household size = 4
- Purpose = Non-work
- Origin TAZ parking = \$0

- Destination TAZ parking = \$0
- Origin TAZ terminal time = 1 minute
- Destination TAZ terminal time = 1 minute

A graphical comparison of a modeled travel time decay curve to the sample travel time decay curve for these attributes is shown in Figure 2. In addition, a plot of difference in model and the sample trip survival probabilities for these attributes is provided in Figure 3.

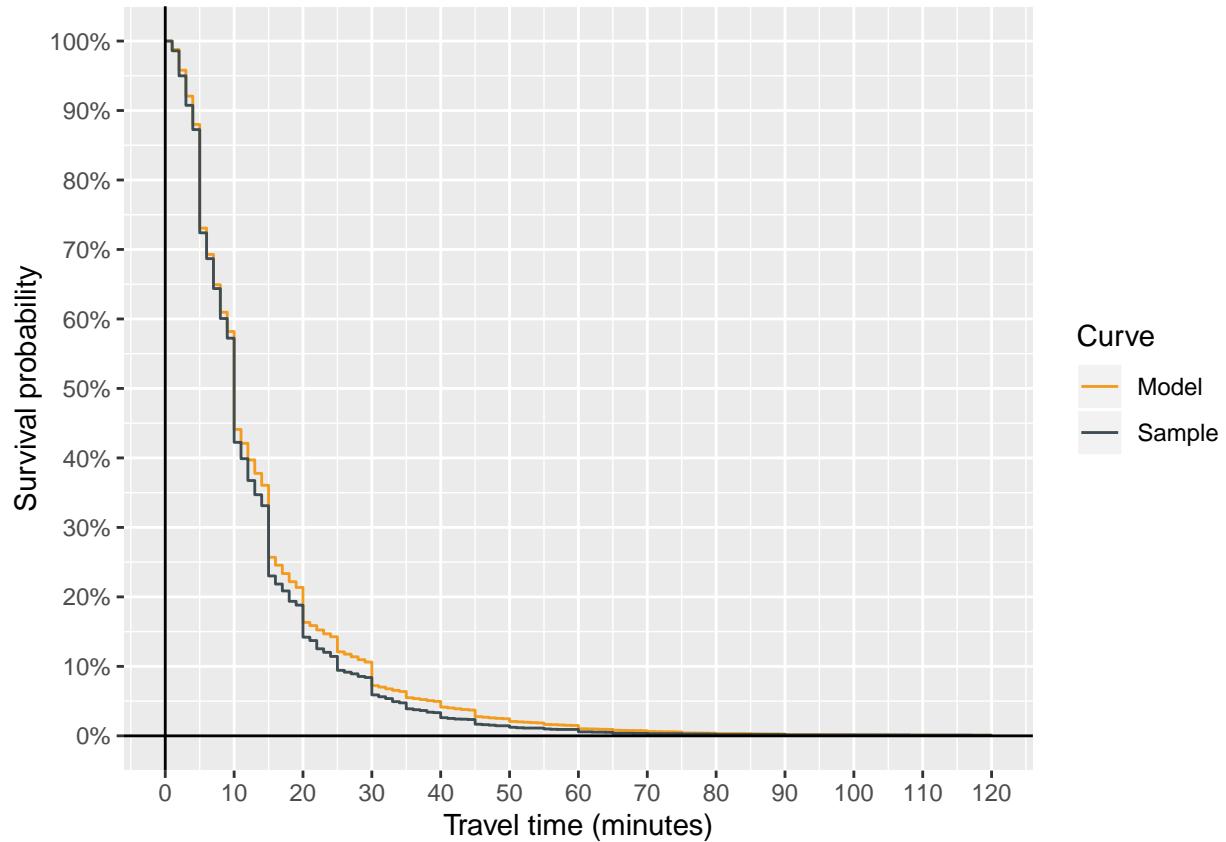


Figure 2: Example comparison of model and sample travel time decay curves for personal vehicle travel

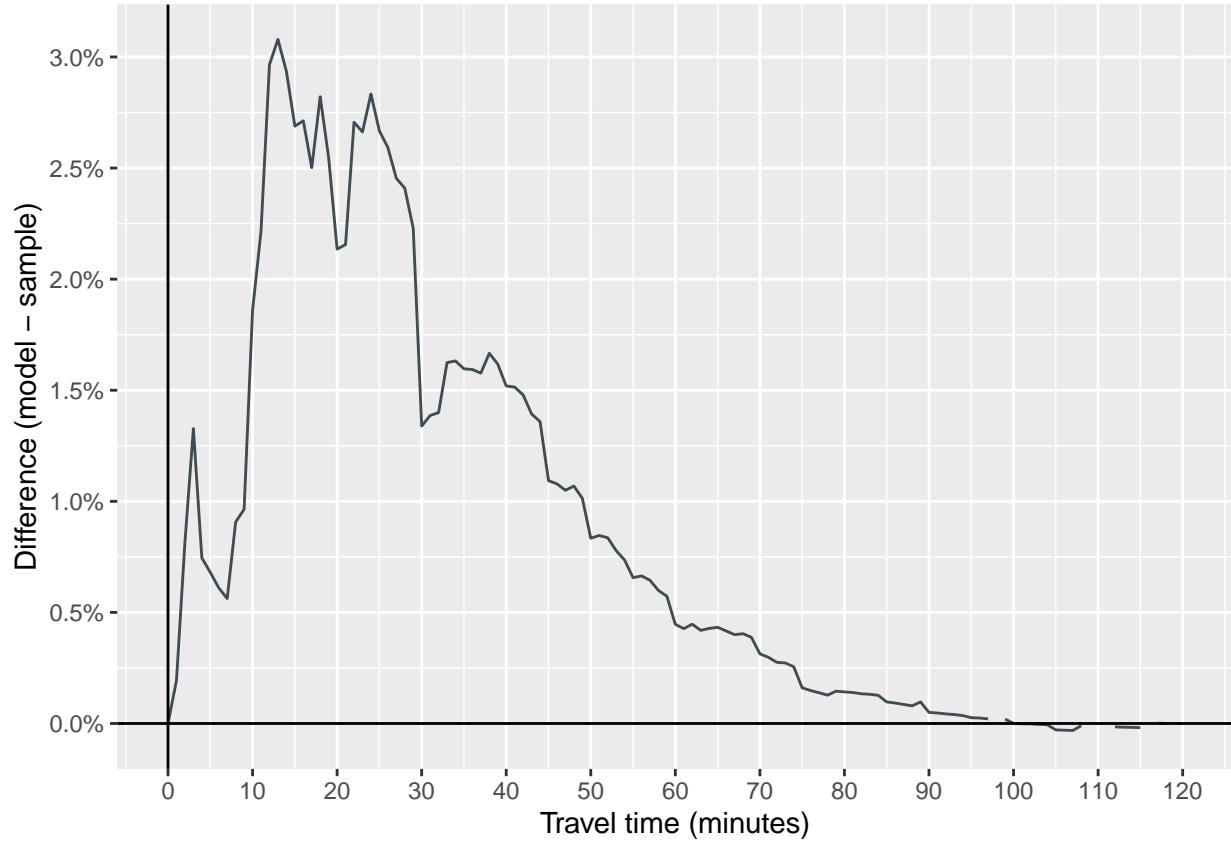


Figure 3: Difference in example model and sample survival probabilities for personal vehicle travel

As evidenced in Figures 2 and 3, the model predicted a slower travel time decay than is reflected in the sample, overpredicting survival probability at nearly every point. However, that the maximum absolute error at any time point is just over 3% indicated that the model reflects the sample well.

4.2 Non-motorized

The model including all candidate covariates showed household size was an insignificant contributor to hazard. When backward selection was used to exclude household size, a statistically valid fit was achieved. The final initial model included only purpose (again, the purpose “home” was used as the reference level for fitting the purpose covariate). The model coefficients are provided in Table 10, while the hazard ratios are provided in Table 11.

Table 10: Coefficients for initial non-motorized model

Covariate	β	$SE(\beta)$	z	$Pr(> z)$
Destination TAZ parking	0.0053	0.0007	7.6080	< 0.0001
Destination TAZ terminal time	-0.1714	0.0094	-18.2873	< 0.0001
Purpose: work	0.3069	0.0235	13.0760	< 0.0001
Purpose: non-work	0.3908	0.0165	23.6941	< 0.0001

Table 11: Hazard ratios (HR) for initial non-motorized model

Covariate	HR	Inverse HR	95% CI for HR
Destination TAZ parking	1.0053	0.9947	[1.0039, 1.0067]
Destination TAZ terminal time	0.8425	1.1869	[0.8272, 0.8581]
Purpose: work	1.3592	0.7357	[1.2981, 1.4232]
Purpose: non-work	1.4782	0.6765	[1.4312, 1.5267]

Schoenfeld Residuals Tests on the covariates revealed time dependence in the non-work purpose, but not for the work purpose. The results of the test are shown in Table 12. A time-dependent coefficients model was explored as a result of present time dependence.

Table 12: Results of Schoenfeld Residuals Tests for initial non-motorized model

Covariate	ρ	χ^2	p
Destination TAZ parking	0.0332	21.8417	< 0.0001
Destination TAZ terminal time	0.0312	19.3586	< 0.0001
Purpose: work	-0.0134	3.8630	0.0494
Purpose: non-work	-0.0548	64.7239	< 0.0001
Global		126.6509	< 0.0001

Exploration of scaled Schoenfeld residuals vs. time plots followed by iterative model fitting and re-checking of the proportional hazards assumption resulted in a time-dependent coefficients model with a single break point at 5 minutes. This model featured no time dependence, and thus this model was used going forward. The model coefficients are provided in Table 13, while the hazard ratios are provided in Table 14.

Table 13: Coefficients for time-dependent non-motorized model

Covariate	Interval	β	SE(β)	z	Pr(> z)
Destination TAZ parking	0-5 minutes	0.0024	0.0010	2.3269	0.02
Destination TAZ parking	> 5 minutes	0.0084	0.0009	9.0147	< 0.0001
Destination TAZ terminal time	0-5 minutes	-0.2198	0.0162	-13.5900	< 0.0001
Destination TAZ terminal time	> 5 minutes	-0.1465	0.0115	-12.7736	< 0.0001
Purpose: work	0-5 minutes	0.3530	0.0371	9.5180	< 0.0001
Purpose: work	> 5 minutes	0.2881	0.0304	9.4703	< 0.0001
Purpose: non-work	0-5 minutes	0.5299	0.0264	20.0958	< 0.0001
Purpose: non-work	> 5 minutes	0.2871	0.0214	13.4093	< 0.0001

Table 14: Hazard ratios (HR) for time-dependent non-motorized model

Covariate	Interval	HR	Inverse HR	95% CI for HR
Destination TAZ parking	0-5 minutes	1.0024	0.9976	[1.0004, 1.0045]
Destination TAZ parking	> 5 minutes	1.0085	0.9916	[1.0066, 1.0103]
Destination TAZ terminal time	0-5 minutes	0.8027	1.2459	[0.7776, 0.8285]
Destination TAZ terminal time	> 5 minutes	0.8637	1.1578	[0.8445, 0.8833]
Purpose: work	0-5 minutes	1.4233	0.7026	[1.3235, 1.5306]
Purpose: work	> 5 minutes	1.3339	0.7497	[1.2567, 1.4159]
Purpose: non-work	0-5 minutes	1.6988	0.5887	[1.6132, 1.7889]
Purpose: non-work	> 5 minutes	1.3325	0.7504	[1.2778, 1.3897]

The plot of deviance residuals for the updated, time-dependent model is shown in Figure 4. Though a few potential outliers are observed in the “trip ended later than expected” category (very negative deviance residuals), they are small in number, and individual exploration revealed no grounds on which to exclude any points as outliers.

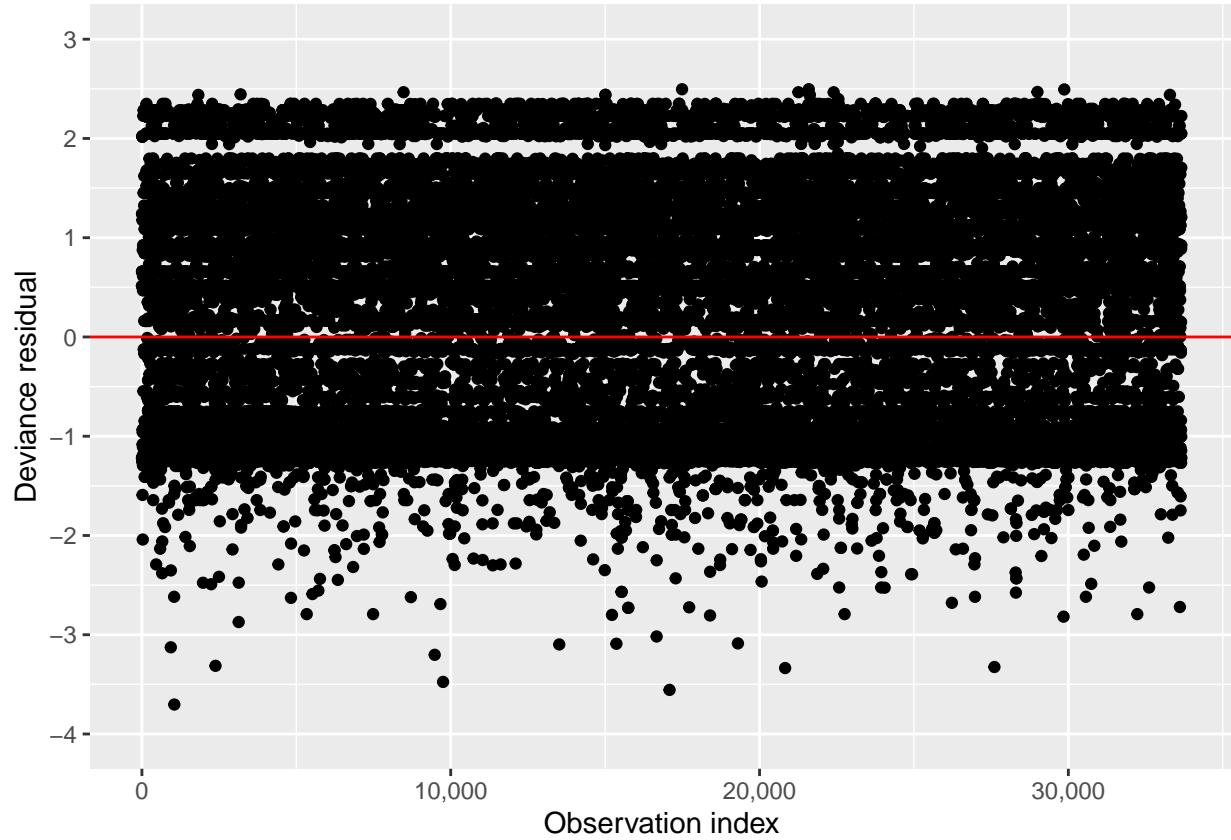


Figure 4: Time dependent non-motorized model deviance residuals

Plots of null Cox model Martingale residuals against destination TAZ parking and destination TAZ terminal time revealed no drastic deviations from linearity, so no functional forms were changed in the time-dependent model. However, like with the personal vehicle model, exact linearity did not appear to be achieved for any of these covariates, indicating better functional forms likely exist for the inclusion of these covariates in

the model. For this analysis, linearity was deemed “close enough” in the context of time and processing limitations.

Having estimated and validated a Cox proportional hazards model for non-motorized travel time decay, the estimated survival curve was compared to the sample curve for the “most common set of conditions”. For non-motorized travel, the most common set of conditions was:

- Destination TAZ parking = \$0
- Destination TAZ terminal time = 0 minutes
- Purpose = Non-work

A graphical comparison of a modeled travel time decay curve to the sample travel time decay curve for these attributes is shown in Figure 5. In addition, a plot of difference in model and the sample trip survival probabilities for these attributes is provided in Figure 6.

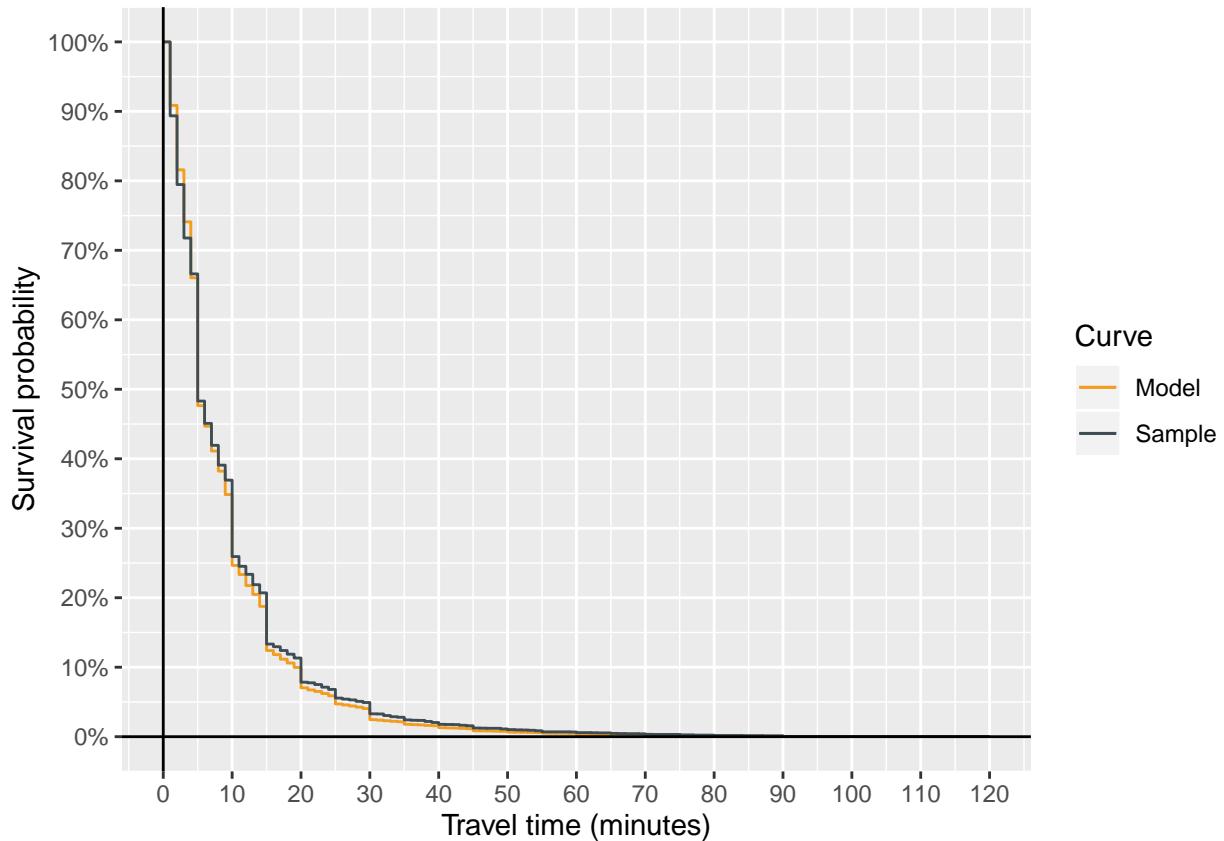


Figure 5: Example comparison of model and sample travel time decay curves for non-motorized travel

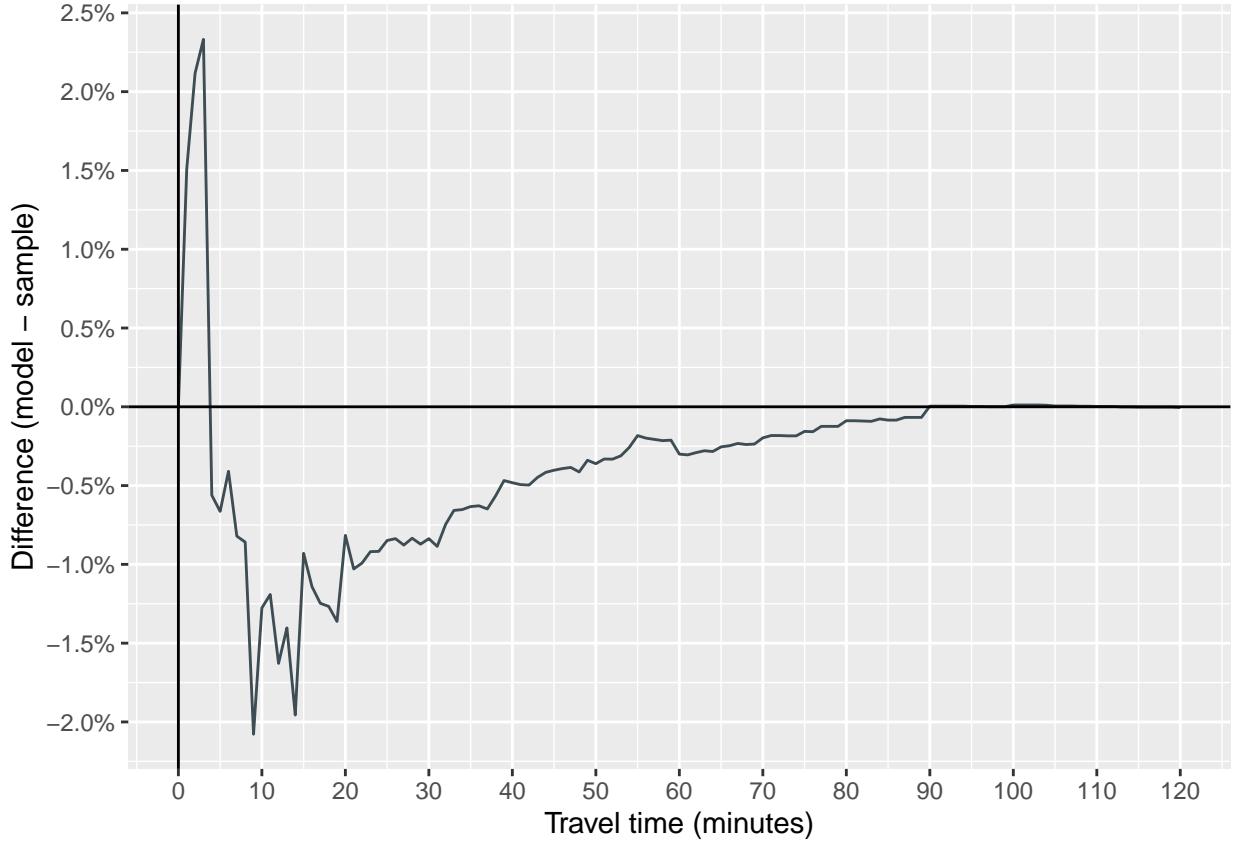


Figure 6: Difference in example model and sample survival probabilities for non-motorized travel

As evidenced in Figures 5 and 6, the model predicted a quicker travel time decay than is reflected in the sample, underpredicting survival probability at most points. However, that the maximum absolute error at any time point is just over 2% indicated that the model reflects the sample well.

4.3 Transit

5 Operationalizing the model

Ultimately, it is hoped that this model would be used to answer the following question: *how would a change in policy impact trip making?* Answering this question relies on comparing the estimated probability density of survival times under varying conditions, and adjusting for natural de-valuing of longer distance trips.

5.1 Deriving a decay factor

Consider two cases: a null case C_0 representing the existing conditions, and an alternative case C_A representing some sort of change in the travel system (e.g. an increase in toll cost). Define a decay factor D_{t,C_0,C_A} as the value of a trip of duration t based on:

1. the duration of the trip, t
2. the change in conditions from C_0 to C_A .

5.1.1 Trip duration component

For the trip duration component, we need to tackle the question *what proportion of people are willing to travel t minutes?* To answer this, consider the population of people travelling more than t minutes under the existing conditions C_0 . Recall that the survival function $S(t) = P(T > t)$, so this population is $(100 \cdot S_{C_0}(t))\%$ of all people. The trip durations for people in this set include $t+1, t+2, \dots, t_{max}$ (because time is discrete in this analysis). Under equal circumstances, we assume that a person whose trip falls in this set would choose to travel the minimum time of the set if given the choice. In other words, all people travelling longer than t would choose to travel $t+1$ if they could. By this logic, $S_{C_0}(t)$ is a measure of “willingness to travel $t+1$ minutes”: $(100 \cdot S_{C_0}(t))\%$ would travel $t+1$ minutes if they could. It follows that $S_{C_0}(t-1)$ gives the proportion of people willing to travel t minutes, which answers the initial question.

5.1.2 Change in conditions component

For the change in conditions component, we need to tackle the question *how does the likelihood of taking a t minute trip change between C_0 and C_A ?* To answer this, we can rely on the formulation of survival curves. Consider the following mathematical equivalencies:

$$\begin{aligned}
 S(t) &= P(T > t) && \text{(definition of survival function)} \\
 &= 1 - P(T \leq t) && \text{(probability rules)} \\
 &= 1 - F(t) && \text{(definition of cdf)} \\
 &= 1 - \int_0^t f(t)dt && \text{(relationship between cdf and pdf)} \\
 &= 1 - \sum_0^t f(t) && \text{(equivalency due to discrete nature of } t\text{)} \\
 &= 1 - \sum_0^t P(T = t) && \text{(definition of pdf)}
 \end{aligned}$$

$f(t) = P(T = t)$ is the value of interest here, because it defines the likelihood of taking an exactly t minute trip. To understand how $f(t)$ can be calculated from survival probabilities, consider the following:

$$F(t) = \sum_0^t f(t)$$

Then: $F(0) = f(0)$

$$= 0$$

$$F(1) = f(0) + f(1)$$

$$= F(0) + f(1)$$

$$\implies \mathbf{f}(\mathbf{1}) = \mathbf{F}(\mathbf{1}) - \mathbf{F}(\mathbf{0})$$

$$F(2) = f(0) + f(1) + f(2)$$

$$= F(1) + f(2)$$

$$\implies \mathbf{f}(\mathbf{2}) = \mathbf{F}(\mathbf{2}) - \mathbf{F}(\mathbf{1})$$

$$F(3) = f(0) + f(1) + f(2) + f(3)$$

$$= F(2) + f(3)$$

$$\implies \mathbf{f}(\mathbf{3}) = \mathbf{F}(\mathbf{3}) - \mathbf{F}(\mathbf{2})$$

⋮

$$F(t) = f(0) + f(1) + \dots + f(t)$$

$$= F(t-1) + f(t)$$

$$\implies \mathbf{f}(\mathbf{t}) = \mathbf{F}(\mathbf{t}) - \mathbf{F}(\mathbf{t}-\mathbf{1})$$

It then follows that:

$$\begin{aligned} f(t) &= F(t) - F(t-1) \\ &= 1 - S(t) - [1 - S(t-1)] \\ &= 1 - S(t) - 1 + S(t-1) \\ &= S(t-1) - S(t) \end{aligned}$$

So, the likelihood of taking an exactly t minute trip $f(t) = S(t-1) - S(t)$. For our cases C_0 and C_A , two unique survival functions $S_{C_0}(t)$ and $S_{C_A}(t)$ will be estimated, resulting in unique estimates $f_{C_0}(t)$ and $f_{C_A}(t)$ for the probability of taking a t minute trip in C_0 and C_A respectively. With these estimates, expected change can be expressed through a density ratio: $\frac{f_{C_A}(t)}{f_{C_0}(t)}$ represents the factor by which the amount of t minute trips should change with a switch from C_0 to C_A , which answers the initial question.

5.1.3 Calculating decay

Let $S_{C_0}(t-1)$ be the proportion of people who would be willing to travel t minutes. Furthermore, let the density ratio $\frac{f_{C_A}(t)}{f_{C_0}(t)}$ be the factor of expected change associated with trips of t minutes. Then we can calculate a decay factor of the following form:

$$D_{t,C_0,C_A} = S_{C_0}(t-1) \cdot \frac{f_{C_A}(t)}{f_{C_0}(t)}$$

D_{t,C_0,C_A} can be interpreted as the proportion of people who would be expected to travel t minutes if given the opportunity in C_A , assuming the willingness to travel reflected by behavior in C_0 .

5.2 A walkthrough example

To exemplify this operationalization, imagine that we want to quantify the value of a 15 minute vehicle trip in the context of increasing tolls. The null case C_0 will assume no tolls. It will be compared to three alternative cases $C_{A_i}, i = 1, 2, 3$, where i represents a \$1, \$2, or \$3 toll, respectively. All other model variables will be held constant. Table 15 details the covariates in each case.

Table 15: Covariates for C_0 and $C_{A_i}, i = 1, 2, 3$

Case	Parking cost	Toll cost	Household size	Purpose	Orig. TAZ parking	Dest. TAZ parking	Orig. TAZ term. time	Dest. TAZ term. time
C_0	\$0	\$0	1	Work	\$0	\$0	0	0
C_{A_1}	\$0	\$1	1	Work	\$0	\$0	0	0
C_{A_2}	\$0	\$2	1	Work	\$0	\$0	0	0
C_{A_3}	\$0	\$3	1	Work	\$0	\$0	0	0

The first step involves calculating the proportion of people willing to travel 15 minutes, or $S_{C_0}(14)$. Using the personal vehicle model, we predict a survival curve using the conditions in C_0 , and extract the survival probability at $t = 14$. We find that $S_{C_0}(14) = 0.644$. A visualization is provided in Figure XXX.

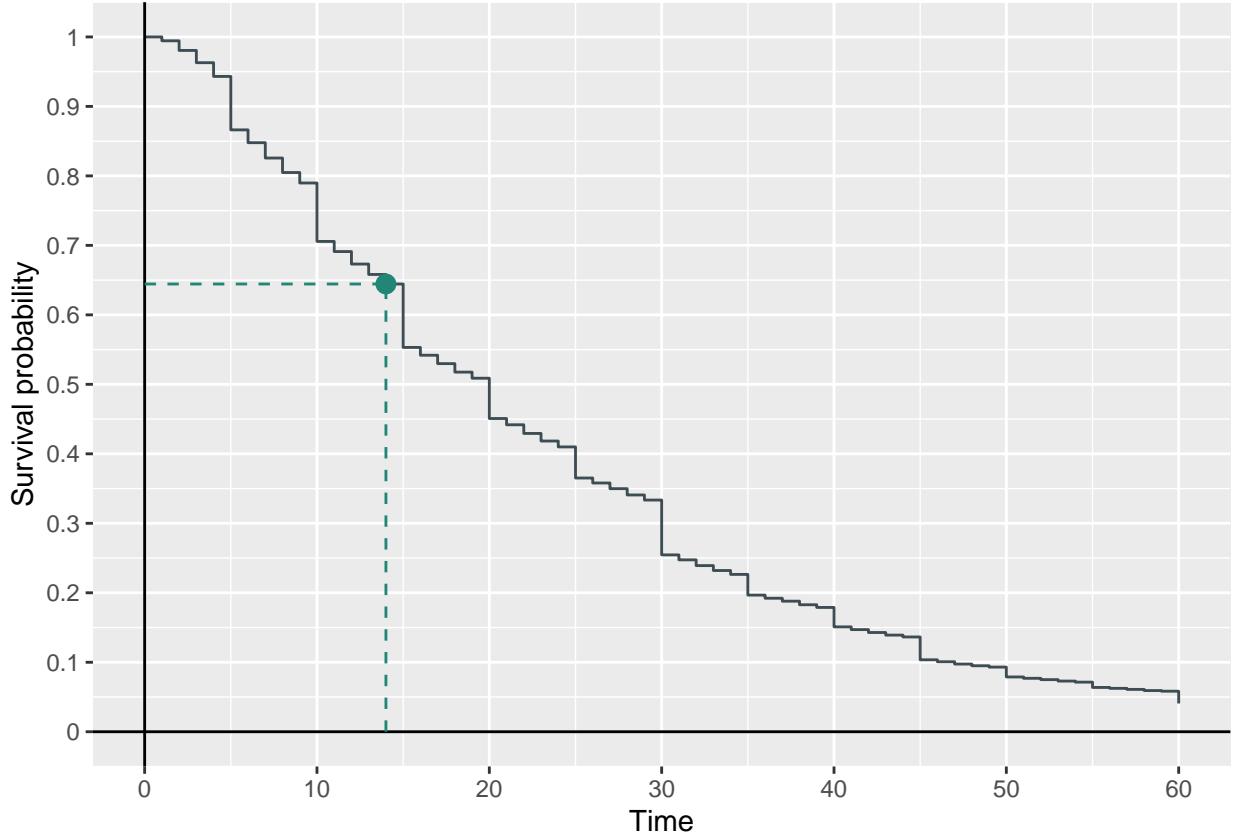


Figure 7: $S_{C_0}(t)$, with $S_{C_0}(14)$ highlighted

Next, the probabilities that a trip lasts 15 minutes must be calculated in each alternative case, and then compared to the same probability in the null case. Using the personal vehicle model, we predict survival curves for C_{A_1} , C_{A_2} , and C_{A_3} , and calculate $f_{C_{A_i}}(15)$ according to the survival probabilities at $t = 14, 15$.

We refer back to the null case survival curve to calculate $f_{C_0}(15)$ in the same manner. Finally, we take the ratio of each $f_{C_{A_i}}(15)$ to $f_{C_0}(15)$. The results are provided in Table XXX. For context, the null and alternative case probability density functions on $t \in [10, 20]$ are shown in Figure XXX.

Table 16: $f_{C_0}(t)$, $f_{C_{A_i}}(t)$, and $\frac{f_{C_{A_i}}(15)}{f_{C_0}(15)}$ for 15 minute vehicle trips

Case	Density	Density ratio (to C_0)
C_0	0.091	1.000
C_{A_1}	0.088	0.968
C_{A_2}	0.062	0.679
C_{A_3}	0.042	0.461

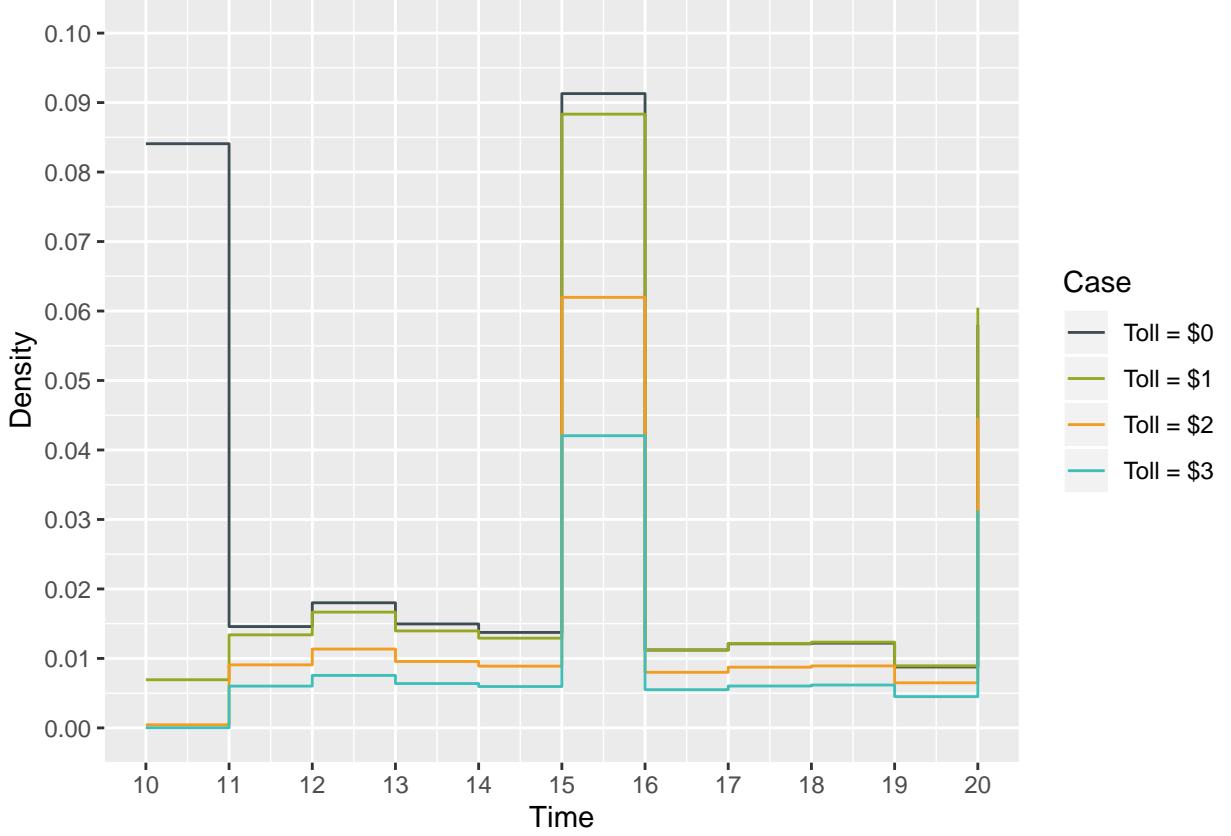


Figure 8: $f_{C_0}(t)$ and $f_{C_{A_i}}(t)$, $i = 1, 2, 3$

Finally, the decay factor $D_{15,C_0,C_{A_i}} = S_{C_0}(14) \cdot \frac{f_{C_{A_i}}(15)}{f_{C_0}(15)}$ is calculated for all three C_{A_i} . The results are provided in Table 18.

Table 17: Calculation of $D_{15,C_0,C_{A_i}}$

Case	Willingness to travel	Density ratio	Decay factor
C_{A_1}	0.644	0.968	0.624
C_{A_2}	0.644	0.679	0.437
C_{A_3}	0.644	0.461	0.297

So, it is learned that, without even considering toll increases, 15 minute trips are de-valued (relative to a less-than-1 minute trip) by a factor of about 0.65 based on time alone. The impact of the toll increase on a 15 minute trip is negligible at \$1 (estimated 0.97 times the 15 minute trips), but is notable at \$2 and \$3 (estimated 0.68 and 0.46 times the 15 minute trips, respectively). It can be concluded that, for the about 65% of travelers willing to take a 15 minute trip, a \$1 toll does not drastically alter accessibility, but that a \$2 or more toll does.

6 Conclusion

A time-dependent Cox proportional hazards model offers an improvement over classical travel time decay modeling methods by allowing for the inclusion of covariate information and quantifying effect sizes associated with these covariates. Furthermore, the use of time-dependent coefficients in these Cox models allows for an interval understanding of covariate effects, which is particularly useful in travel time since trips are often thought of in 5 or 10 minute intervals. The Cox models fit in this analysis indicate that, depending on the mode of choice, trip purpose, trip cost, and household size may all be significant contributors to trip hazard rate and, consequently, trip survival time. These covariates represent a suite of information that should be readily available for use of these modeling in a predictive or simulative context.

Future models should seek to explore additional relevant covariate information (such as terminal time) and/or potential interactions between covariates. They also may seek the following modeling improvements, which were not addressed in this analysis due to processing and time constraints.

1. *Use of a smoothed function for time-dependent coefficients, as opposed to a step function:* Rather than estimating coefficients in discrete time intervals, coefficients could be estimated using a smoothed function of time. For example, in the case of monotonically decreasing effect size with time, a linear function could be valid. This could improve accuracy in estimating effect sizes
2. *Estimation of proper functional form for continuous covariates:* Rather than assuming a basic linear form for continuous covariates, more representative functional forms could be estimated using smoothing techniques, for example. This would improve understanding of the relationship between the covariates and hazard, as well as likely improve predictive accuracy.