CIVIL-557
Car-following models: Specification and estimation

Evangelos Paschalidis

Transport and Mobility Laboratory (TRANSP-OR)
École Polytechnique Fédérale de Lausanne (EPFL)

Lecture overview

- 08:15 - 09:00 Fundamentals of driving behaviour models
- 09:15 - 10:00 The car-following model
- 10:15 - 12:00 The car-following model (Lab session)
    - Model specification - estimation
    - Model interpretation

Fundamentals of driving behaviour models

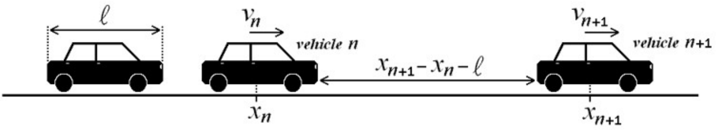
- What are the driving behaviour models
- Applications and usefulness
- Types of driving behaviour models
    - Car-following
    - Gap-acceptance
    - Lane-changing
- Relevance to statistics and statistical modelling

Car-following models

- Types of car-following models
- Data requirements
- Model specification and estimation

Car-following models

- Car-following (CF) models else follow the leader models – acceleration responses with respect to the lead vehicle behaviour
- Possibly the most studied type of driving behaviour models



- One of the most common types of car following models is the stimulus response models defined as:

$$\text{response (acceleration)} = \text{sensitivity x stimulus}$$

Car-following models

Some other types of car-following models are:

- *Desired measures models*: drivers follow the leader based on some desired metric (e.g. desired spacing distance) [Examples: Helly's model, Intelligent Driver Model (IDM)]
- *Collision avoidance models*: Reaction of drivers on relative spacing distance and not on relative speed [Example: Gipps model]
- *Optimal velocity models*: Drivers aim at an optimal speed which they perceive as safe [Example: Bando's model]
- *Perceptual thresholds models*: Areas within a range of relative speed and relative distance where a driver reacts [Example: Wiedemann's model]

Stimulus-response models

- The first stimulus-response models [else known as the GM model] were developed at the General Motors research laboratories (late 50's - early 60's)
- A typical functional form of the model is

$$\alpha_n(t) = \alpha \frac{V_n(t)^\beta}{\Delta X_n(t-\tau_n)^\gamma} \Delta V_n(t-\tau_n)^\lambda$$

where

- $\alpha_n(t)$ is acceleration of individual n at time t
- $V_n(t)$ is speed of individual n at time t
- $\Delta X_n(t-\tau_n)$ is the space headway of individual n with the lead vehicle at time $t-\tau$ where τ_n is the reaction time of individual n
- $\Delta V_n(t-\tau_n)$ is the relative speed of individual n with the lead vehicle at time $t-\tau_n$
- α, β, γ and λ are parameters to be estimated

Desired measures models

- Address the limitation of GM model that for two vehicles travelling at the same speed, any spacing between them is acceptable
- Helly's (1959) model is a typical example

$$\alpha_n(t) = \alpha_1 \Delta V_n(t-\tau_n) + \alpha_2[\Delta X_n(t-\tau_n) - \overline{\Delta X}_n(t)]$$

$$\overline{\Delta X}_n(t) = \beta_1 + \beta_2 V_n(t-\tau)$$

where

- $\alpha_n(t)$ is acceleration of individual n at time t
- $V_n(t)$ is speed of individual n at time t
- $\Delta X_n(t-\tau)$ is the space headway of individual n with the lead vehicle at time $t-\tau$ where τ is the reaction time
- $\Delta V_n(t-\tau)$ is the relative speed of individual n with the lead vehicle at time $t-\tau$
- $\overline{\Delta X}_n(t)$ is the desired space headway
- $\alpha_1, \alpha_2, \beta_1$ and β_2 are parameters to be estimated

Desired measures models

The intelligent driver model (IDM) is another example of a desired measures model (Treiber et al., 2000):

$$\alpha_n(t) = \alpha_{max}^{(n)} \left[1 - \left(\frac{V_n(t)}{\overline{V}_n(t)} \right)^{\beta} - \left(\frac{\overline{S}_n(t)}{S_n(t)} \right)^2 \right]$$

where

- $\alpha_n(t)$ is acceleration of individual n at time t
- $\alpha_{max}^{(n)}$ is the maximum acceleration of individual n [parameter to be estimated]
- $V_n(t)$ is speed of individual n at time t
- $\overline{V}_n(t)$ is the desired speed speed of individual n at time t [parameter to be estimated]
- $S_n(t)$ is the spacing distance of individual n with the lead vehicle at time t
- $\overline{S}_n(t)$ is the desired spacing distance of individual n with the lead vehicle at time t [to be estimated]
- β is a parameter (usually value fixed to 4)

Desired measures models

The desired spacing distance in the IDM model is specified as (simplified version):

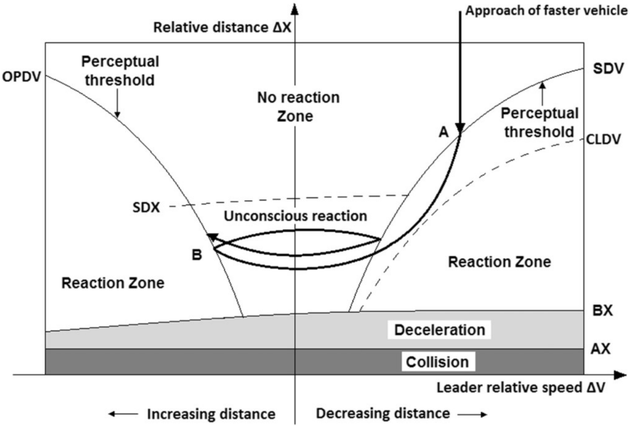
$$\overline{S}_n(t) = S_0^{(n)} + S_1^{(n)} \sqrt{\frac{V_t(t)}{\overline{V}_n(t)}} + V_t(t)\overline{T}_n(t)$$

where

- S_0^n us the minimum spacing distance at the standstill situation
- S_1^n us the minimum spacing distance at the standstill situation interacting with speed
- $\overline{V}_n(t)$ is the desired speed speed of individual n at time t [parameter to be estimated]
- $\overline{T}_n(t)$ is the desired time headway at time t

Perceptual thresholds models

- The most common example is Wiedemann's model (1974)
- There are perceptual zones depending on relative speed and distance where drivers react or unconsciously follow



Data needs and format

First, let's consider what's the most common variables we need

- Acceleration (serves as dependent variable)
- Speed
- Speed of lead vehicle
- Distance (space or time) with the lead vehicle
- Traffic density (less common)

– These variables are sufficient for most car-following models but additional are required for other models such as lane-changing or gap-acceptance

Data needs and format

Not all information that we need is always there (default NGSIM I-80 data structure)

Vehicle_ID	Frame_ID	Lane_ID	Local_Y	Mean_Speed	Mean_Accel	Vehicle_Le	Vehicle_Class_ID	Follower_ID	Leader_ID
46	861	6	1417.646	37.86434	-0.0396	16.80184	2	45	-1
46	862	6	1421.431	37.85774	-0.06604	16.80184	2	45	-1
46	863	6	1425.216	37.84672	-0.11014	16.80184	2	45	-1
46	864	6	1428.999	37.829	-0.17707	16.80184	2	45	-1
46	865	6	1432.779	37.80282	-0.26194	16.80184	2	45	-1
46	866	6	1436.556	37.77031	-0.32493	16.80184	2	45	-1
46	867	6	1440.33	37.74059	-0.29731	16.80184	2	45	-1
46	868	6	1444.103	37.72546	-0.15148	16.80184	2	45	-1
46	869	6	1447.876	37.72953	0.04072	16.80184	2	45	-1
46	870	6	1451.65	37.74449	0.1497	16.80184	2	45	-1
46	871	6	1455.426	37.75358	0.09081	16.80184	2	45	-1
46	872	6	1459.2	37.74016	-0.13402	16.80184	2	45	-1
46	873	6	1462.969	37.69137	-0.48786	16.80184	2	45	-1
46	874	6	1466.729	37.59826	-0.93107	16.80184	2	45	-1
46	875	6	1470.475	37.46122	-1.37034	16.80184	2	45	-1
46	876	6	1474.205	37.29774	-1.63481	16.80184	2	45	-1
47	549	1	165.965	37.01401	0	14.90158	2	-1	39
47	550	1	169.6863	37.21319	1.9919	14.90158	2	-1	39
47	551	1	173.4069	37.20571	-0.07474	14.90158	2	-1	39
47	552	1	177.1261	37.19216	-0.13563	14.90158	2	-1	39
47	553	1	180.8435	37.17398	-0.18179	14.90158	2	-1	39
47	554	1	184.5591	37.15568	-0.18301	14.90158	2	-1	39
47	555	1	188.2729	37.13858	-0.17087	14.90158	2	-1	39
47	556	1	191.9849	37.12021	-0.18369	14.90158	2	-1	39
47	557	1	195.6945	37.09557	-0.24652	14.90158	2	-1	39
47	558	1	199.4002	37.05679	-0.38789	14.90158	2	-1	39
47	559	1	203.0993	36.99134	-0.6543	14.90158	2	-1	39
47	560	1	206.7872	36.87914	-1.12228	14.90158	2	-1	39

Data needs and format

Intense data process is usually required...

ID	Time	Position	Length	Width	Type	Speed	Acceleratic	Lane	Leader	Follower	Space_headway	Time_headway	Acceleratic	Speed_lead	Position_le	Density	distanceTo
4	26	83.0336	4.084802	-9999.99	2	6.84778	-2.26641	5	21	27	15.25370076	2.227539529	0	0	98.2873	7	361.6202
4	27	87.75906	4.084802	-9999.99	2	3.29583	-1.34933	5	21	27	10.5282431	3.194413438	0	0	98.2873	6	356.8947
4	28	90.96072	4.084802	-9999.99	2	2.820909	-2.04014	5	21	27	7.399970976	2.623257824	1.049439	0.34478976	98.36069	6	353.6931
4	29	92.04724	4.084802	-9999.99	2	0.276981	0.345661	5	21	27	6.925830192	25.00472087	0.67821	0.796451544	98.97307	6	352.6065
4	30	92.60182	4.084802	-9999.99	2	0.634731	0.36587	5	21	27	7.371548376	11.61366179	1.018739	1.222229712	99.97337	6	352.0519
4	31	93.58288	4.084802	-9999.99	2	0.99491	-0.10506	5	21	27	7.74211812	7.781728112	-0.12618	1.245559104	101.325	8	351.0709
4	32	94.93359	4.084802	-9999.99	2	1.516721	0.07495	5	21	27	7.654741104	5.046899994	-0.8841	1.178749992	102.5883	8	349.7202
4	33	96.49206	4.084802	-9999.99	2	1.648489	0.32919	5	21	27	6.947029032	4.214178606	0.381939	0.826760856	103.4391	8	348.1617
4	34	97.68815	4.084802	-9999.99	2	0.75071	-0.40722	5	21	27	6.516840408	8.680900218	-0.05167	0.695160408	104.205	7	346.9656
4	35	98.41664	4.084802	-9999.99	2	0.754371	0.049579	5	21	27	6.678119232	8.85256791	-0.89062	0.814910232	105.0948	8	346.2371
4	36	99.17668	4.084802	-9999.99	2	0.757169	-0.02056	5	21	27	6.567028776	8.67313568	0.859201	0.821688984	105.7437	8	345.4771
4	37	99.84506	4.084802	-9999.99	2	0.55904	-0.1709	5	21	27	7.378378944	13.19830764	0.531111	1.735689696	107.2234	8	344.8087
4	38	100.3861	4.084802	-9999.99	2	0.55216	0.08143	5	21	27	8.980928952	16.26507135	0.426241	2.326011096	109.367	9	344.2677
4	39	101.2736	4.084802	-9999.99	2	1.366681	1.35711	5	21	27	10.99218833	8.042983055	0.57186	3.364611	112.2658	9	343.3801
4	40	103.2714	4.084802	-9999.99	2	2.34176	0.528529	5	21	27	12.91753068	5.516163083	0.82118	4.383148968	116.1889	9	341.3824
4	41	106.4301	4.084802	-9999.99	2	3.69478	0.76741	5	21	27	14.65547894	3.96653682	0.512329	5.178289872	121.0855	9	338.2237
4	42	111.0489	4.084802	-9999.99	2	5.270032	0.511521	5	21	27	15.52339999	2.945599021	0.345341	5.902461144	126.5723	9	333.6048
4	43	116.473	4.084802	-9999.99	2	5.60665	0.57955	5	21	27	16.25228957	2.898752455	0.3367	6.575319336	132.7253	9	328.1808
4	44	122.5318	4.084802	-9999.99	2	6.575152	0.88029	5	21	27	16.15559177	2.457067535	-1.57026	5.29724112	138.6874	9	322.122
4	45	129.228	4.084802	-9999.99	2	6.81378	1.109249	5	21	27	14.51480155	2.130212946	0.54543	5.1356514	143.7428	8	315.4258
4	46	136.2907	4.084802	-9999.99	2	6.500869	-1.46755	5	21	27	12.821031	1.972202673	0.652799	5.564050848	149.1117	8	308.3631
4	47	142.8272	4.084802	-9999.99	2	6.857061	0.659971	5	21	27	12.17461116	1.775485266	0.375501	6.048460344	155.0018	7	301.8266
4	48	149.6461	4.084802	-9999.99	2	6.732261	-0.05924	5	21	27	11.81505079	1.754990046	-0.00769	6.657170328	161.4611	7	295.0077
4	49	156.5059	4.084802	-9999.99	2	6.98751	0.12399	5	21	27	11.61510199	1.662266357	-0.08365	6.67265112	168.121	7	288.1478
7	24	82.64016	5.426001	-9999.99	2	6.38645	-2.36717	6	5	41	11.69280161	1.830876556	-0.21628	6.101510784	94.33296	6	364.2222
7	25	88.17163	5.426001	-9999.99	2	5.14642	-0.52017	6	5	41	11.56951915	2.248071317	-1.00729	4.666399608	99.74115	5	358.6907
7	26	92.42863	5.426001	-9999.99	2	3.458279	-1.66266	6	5	41	11.20113787	3.238934044	-2.07837	2.897309928	103.6298	4	354.4337
7	27	95.2192	5.426001	-9999.99	2	2.56253	-0.21566	6	5	41	10.97755183	4.283872848	-0.06621	2.503669872	106.1968	4	351.6432

Let's consider a stimulus - response model defined as:

$$response = stimulus \text{ x } sensitivity$$

or

$$\alpha_n^g(t) = s[X_n^g(t-\tau_n)]f[\Delta V_n(t-\tau_n)] + \epsilon_n^g(t)$$

where

- $g \in$ (*acceleration, deceleration*)
- $s[X_n^g(t-\tau_n)]$: sensitivity to stimulus as a function of $X_n^g(t-\tau_n)$
- $X_n^g(t-\tau_n)$: explanatory variables observed at time $t-\tau$
- $f[\Delta V_n(t-\tau_n)]$: stimulus (usually a function of relative speed)
- $\Delta V_n(t-\tau_n)$: relative speed (else $V_n^{Lead}(t-\tau) - V_n(t-\tau)$)
- $\epsilon_n^g(t)$: random disturbance term associated with car-following acceleration/deceleration of individual n at time t (normally distributed, $N \sim [0, \sigma^2]$)

Model specification

- The GM model is a stimulous - response model
- Based on the specifications we have defined we have:

$$s[X_n^g(t-\tau_n)] = \alpha^g \frac{V_n(t)^{\beta^g}}{\Delta X_n(t-\tau_n)^{\gamma^g}}$$

and

$$f[\Delta V_n(t-\tau_n)] = |\Delta V_n(t-\tau_n)|^{\lambda^g}$$

As we mentioned, $\alpha^g, \beta^g, \gamma^g$ and λ^g are parameters to be estimated

⚠ Pay attention to the g supercript in the notation of the parameters!! This indicates that different parameters need to be estimated in acceleration and deceleration conditions.

⚠ This acceleration-deceleration assymetry applies only in this version of the GM model and not in desired measures or other car-following models.

Model estimation

But how can we find the values of the unknown α, β, γ and λ parameters?

In the previous lecture we discussed the maximum likelihood estimation

For continuous variables, under the assumption of a normally distributed disturbance [as we also defined in the GM model] we get:

$$f(Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5(\frac{Y-\mu}{\sigma})^2}$$

A shorter notation is:

$$f(Y) = \frac{1}{\sigma}\phi\left(\frac{Y-\mu}{\sigma}\right)$$

where $\phi()$ represents the density function of a standard normal distribution ($N \sim [0,1]$)

Model estimation

Time to extend our knowledge on the GM model estimation...

The steps to follow are:

1. Specify stimulus and sensitivity for acceleration
2. Specify stimulus and sensitivity for deceleration
3. Specify the density function for acceleration
4. Specify the density function for deceleration
5. Set an acceleration/deceleration rule (i.e. the conditions when to consider the acceleration or deceleration density function
6. Write down the likelihood function for each individual n
7. Estimate the the sum of the log-likelihood of all individuals

We have already covered steps 1-4 in model specification.

Model estimation

The density function in acceleration regime is:

$$f(\alpha_n^{acc}(t)|\tau_n) = \frac{1}{\sigma_{\epsilon^{acc}}}\phi\left(\frac{\alpha_n^{acc}(t)-s[X_n^{acc}(t-\tau_n)]f[\Delta V_n(t-\tau_n)]}{\sigma_{\epsilon^{acc}}}\right)$$

The density function in deceleration regime is:

$$f(\alpha_n^{dec}(t)|\tau_n) = \frac{1}{\sigma_{\epsilon^{dec}}}\phi\left(\frac{\alpha_n^{dec}(t)-s[X_n^{dec}(t-\tau_n)]f[\Delta V_n(t-\tau_n)]}{\sigma_{\epsilon^{dec}}}\right)$$

The assumption in this GM model implementation is that a driver will accelerate if relative speed with the lead vehicle is positive (lead vehicle has higher speed compared to following vehicle).

The total acceleration probability of one observation is given by:

$$f(\alpha_n(t)|\tau_n) = f(\alpha_n^{acc}(t)|\tau_n)^{\Delta V(t-\tau_n)\geq 0} f(\alpha_n^{dec}(t)|\tau_n)^{1-(\Delta V(t-\tau_n)\geq 0)}$$

Model estimation

The total likelihood of the observations of an individual driver n is given by the joint density function:

$$f(\alpha_n(1), \alpha_n(2), ..., \alpha_n(T) | \tau_n) = \prod_{t=1}^{T} f(\alpha_n(t) | \tau_n)$$

- The joint density is conditional on reaction time
- How can we deal with reaction time?
  - Assume a deterministic fixed value for all drivers
  - Introduce heterogeneity in reaction time (next lecture)

Model estimation

For now let's assume that reaction time is known and identical for all drivers. Now, the likelihood function is not conditional on the reaction time...

$$f(\alpha_n(1), \alpha_n(2), ..., \alpha_n(T)) = \prod_{t=1}^{T} f(\alpha_n(t))$$

... the log-likelihood function (for all drivers n) is given by:

$$LL = \sum_{n=1}^{N} ln[f(\alpha_n(1), \alpha_n(2), ..., \alpha_n(T))]$$

... and this is it!

Results

We check for two things when evaluating the parameter estimates:

- The sign of parameter is consistent with our expectations
- The parameter is significant

Our expectations for the parameter estimates are...

Parameter	Interpretation	Expected sign	
		Acceleration regime	Deceleration regime
α	Acceleration (deceleration) constant	+	-
β	Speed	-	+
γ	Space headway	+ / -	+ (mostly)
λ	Relative speed (absolute)	+	+

...we will see why during the lab session.

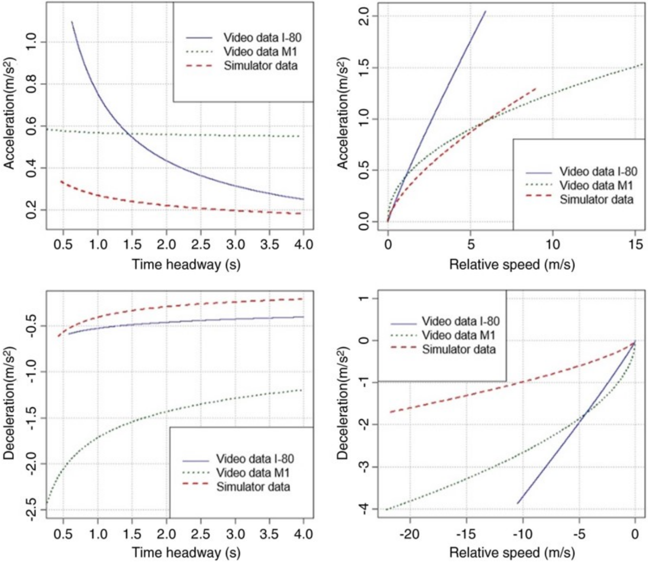
Example

Table 3. Models parameter estimates, t-test of individual parameter equivalence, and TTS results

Variable	Driving simulator data (Model 1) Parameter estimate	Driving simulator data (Model 1) Robust t-ratio	Video trajectory data—I80 (Model 2) Parameter estimate	Video trajectory data—I80 (Model 2) Robust t-ratio	Video trajectory data—M1 (Model 3) Parameter estimate	Video trajectory data—M1 (Model 3) Robust t-ratio
Reaction time distribution						
μ_t	0.664	14.66	−0.3973	−16.42	0.6204	5.72
σ_t	0.3536	2.69	0.3257	66.48	0.6517	4.63
Car-following acceleration						
Constant	0.3506	6.96	0.8304	13.77	0.4283	4.59
Time headway (s)	0.2856	1.85	0.792	8.99	0.0218	0.19
Relative speed (m/s)	0.6787	9.71	0.8982	20.58	0.4694	1.37
σ^{acc}	0.3367	25.26	0.7318	76.71	0.6976	9.07
Car-following deceleration						
Constant	−0.255	−5.47	−0.5128	−16.03	−0.9206	−25.25
Time headway (s)	0.4798	2.66	0.1941	2.36	0.2609	8.67
Relative speed (m/s)	0.7043	9.38	0.928	25.98	0.5195	18.91
σ^{dec}	0.6893	16.61	0.8007	75.16	0.7545	42.55
	LL = −5,610.845		LL = −17,240.88		LL = −3,857.24	
	$\rho^2 = 0.320$		$\rho^2 = 0.138$		$\rho^2 = 0.463$	
	Adj. $\rho^2 = 0.319$		Adj. $\rho^2 = 0.137$		Adj. $\rho^2 = 0.461$	
	obs = 7,191		obs = 14,826		obs = 3,302	

Note: LL= log-likelihood; Adj = adjusted; and obs = observations.

Example


Video data I-80
Video data M1
Simulator data
Acceleration(m/s²)
Time headway (s)
Relative speed (m/s)
Deceleration(m/s²)
0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
0 5 10 15
-20 -15 -10 -5 0

Model interpretation - Steps

- Reminder - Interpretation of parameters in regression type models: How much the dependent variable (DV) Y changes assuming 1-unit change in the independent variable (IV) X_i all else being equal (i.e. the values of all remaining independent variables X_j stays constant)
- In practice: We investigate the impact of the independent variables on the dependent variable one at a time.
- In linear regression: For 1-unit change in the DV, the IV changes by β. What about non-linear models?
- Simplest approach for non-linear models: Sensitivity analysis

Model interpretation - Sensitivity analysis

1. We estimate the model and obtain parameter estimates (results).
2. We decide the IV X_i to investigate.
3. We find the *min* and *max* values of X_i in the date and generate N values (X_{i1},X_{i2},...,X_{iN}) within this range.
4. For the rest X_j independent variables, we give a value equal to their sample average (or median).
5. For each $X_i n$, we compute the expected value of the DV using our model (combining the data and the results from step 1).
6. We plot the results to illustrate the impact of X_i on Y.

Some comments...

- Maximum likelihood estimation can be technically used for other (but not all) car-following model specifications as well (e.g. Helly's, IDM etc.)
- If we want to consider some other car-following model, we have have to fix some of the parameter values because not all models are identified (fixed values can be taken from existing literature)
- The GM model specification does not require fixing any parameter and most times can be estimated relatively smoothly most of the times
- Essentially, estimating car-following models is the same process as estimating a variation of a *linear regression model*

Questions??

Solutions and comments on last week's examples

Crowdshipping logit model

Task: Add time of day as an explanatory (dummy) variable

Solution: Time of the day must be added as a categorical variable. That is three parameters must be estimated for three categories and one is left out as reference (the choice is irrelevant to the final model fit)

Let's also compare if a model with time of day fits the data significantly better than a model without it!

Solutions and comments on last week's examples

Step1: Add the new parameters in the starting values vector

Define parameters and starting values

Ultimately, we want to estimate the value of some parameters that maximise the likelihood of our observations of the dependent variable.

Before starting the estimation process, we need to define some starting values for the parameters to be estimated. The typecal starting value is zero.

```
In [18]: betas_start = {"asc": 0, "b_tt": 0, "b_size":0, "b_fragile":0, "b_reward":0,
                        "b_morning":0,"b_afternoon":0,"b_evening":0}
```

Load old parameter estimates results

Step2: Edit utility function

```
# We need to start by defining the utility functions
# Please make sure that you are using the same names for the parameters as those defined in 'betas_start'

U_accept = (asc + b_tt*trip_duration +
        b_size*Size + b_fragile*Fragile + b_reward*Compensation+
           b_morning*(time_of_day==1) + b_afternoon*(time_of_day==2) + b_evening*(time_of_day==3))
U_reject = 0
```

Solutions and comments on last week's examples

Results

Out[25]:

	Parameter	Estimate	s.e.	t-ratio0	p-value	Rob s.e.	Rob t-ratio0	Rob p-value
0	asc	-1.369	0.116	-11.768	0.000	0.116	-11.786	0.000
1	b_tt	-0.004	0.002	-1.775	0.076	0.002	-1.778	0.075
2	b_size	1.148	0.148	7.749	0.000	0.148	7.755	0.000
3	b_fragile	0.993	0.048	20.499	0.000	0.049	20.441	0.000
4	b_reward	0.028	0.005	5.639	0.000	0.005	5.648	0.000
5	b_morning	0.670	0.073	9.230	0.000	0.073	9.211	0.000
6	b_afternoon	1.381	0.065	21.199	0.000	0.065	21.168	0.000
7	b_evening	1.087	0.064	16.988	0.000	0.064	17.029	0.000

- What's the interpretation of the time of the day variable?
- Is the new model fit significantly better than the model without time of the day included?

Solutions and comments on last week's examples

Does including time of the day significantly improve model fit?

- The results related to time of day suggest a positive impact on utility compared to the reference group (night)
- The significance levels obtained for the time of day referes to the comparison with the reference category only (not e.g between morning and afternoon or any other combination)
- We can evaluate the improvement of model fit (if any) after the addition of time of day using the likelihood ration test as:
  $LR = -2(LL^{restricted} - LL^{unrestricted})$
  where *restricted* is the initial model and *unrestricted* is the model after the addition of time of the day

Solutions and comments on last week's examples

In our example

$LR = -2(LL^{restricted} - LL^{unrestricted}) = -2(-6434.891 - [-6177.035]) =$
515.712

We must evaluate this result for 3 degrees of freedom at the 0.05 level of significance
– We find the critical values in the χ^2-distribution table If LR $> \chi^2$ critical

value then the addition of the new variable is significantly improving model fit In our case this value is $515.712 > 7.81$ hence, model fit is significantly

improved

⚠ LR test can be only used if the old model is a simplified version of the new model. E.g. if all parameters related to time of day were estimated with 0 value then the two models would have identical specification.

Solutions and comments on last week's examples

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21

Solutions and comments on last week's examples

Acceleration regression model

Task 1: Create a relative speed variable by taking $speed\_lead - speed$ *(not in the data)*

Task 2: Linear in parameters model

Solutions and comments on last week's examples

Add distance in the list of variables

Variable definition

We need to define the variables (as numpy arrays) that we will use in our model.

- The arrays can have any name but it is more convenient to use the same name as in the data set.

In [35]:
```
# Example variable_name = np.array(data['variable_name']).reshape(-1, 1)

Distance = np.array(data['Distance']).reshape(-1, 1)
Speed = np.array(data['Speed']).reshape(-1, 1)
Speed_lead = np.array(data['Speed_lead']).reshape(-1, 1)
Acceleration = np.array(data['Acceleration']).reshape(-1, 1)
```

Add parameters for distance and relative speed. Remove parameter of lead vehicle speed

Define parameters and starting values

Ultimately, we want to estimate the value of some parameters that maximise the likelihood of our observations of the dependent variable.

Before starting the estimation process, we need to define some starting values for the parameters to be estimated.

- The starting values are usually zeroes
- When a parameter is included as a denominator, the starting value cannot be 0 for computational reasons.
- However, since we estimate the log of sigma, our starting value can be zero since exp(sigma) can never be absolute zero.

In [39]:
```
betas_start = {"beta0": 0, "beta_speed": 0, "beta_relative_speed": 0, "b_distance":0, "sigma": 0}
```

Load old parameter estimates results

Solutions and comments on last week's examples

Edit the model formula by adding the new explanatory variables

```
# Then we need to define the main model specification
fi = beta0 + beta_speed * Speed + beta_relative_speed * (Speed_lead-Speed) + b_distance*Distance
```

Results

Out[46]:

	Parameter	Estimate	s.e.	t-ratio0	p-value	Rob s.e.	Rob t-ratio0	Rob p-value
0	beta0	0.112	0.024	4.619	0.0	0.023	4.800	0.0
1	beta_speed	-0.301	0.007	-45.935	0.0	0.006	-48.540	0.0
2	beta_relative_speed	0.028	0.005	6.060	0.0	0.005	5.936	0.0
3	b_distance	0.010	0.000	21.595	0.0	0.000	21.914	0.0
4	sigma	-1.184	0.022	-52.955	0.0	0.021	-56.300	0.0

It's your turn now..

Lab session:

- Specification of a basic stimulus-response car-following model
- Examine parameter estimates
- Plot results to get some basic understanding of the independent variables' impact