Published in final edited form as:
Accid Anal Prev. 2010 May ; 42(3): 891–897. doi:10.1016/j.aap.2009.04.022.

Modeling Lateral Control in Driving Studies

Jeffrey D. Dawson, ScD[a], **Joseph E. Cavanaugh, PhD**[a], **K. D. Zamba, PhD**[a], and **Matthew Rizzo, MD**[b,c,d]

Jeffrey D. Dawson: jeffrey-dawson@uiowa.edu; Joseph E. Cavanaugh: joe-cavanaugh@uiowa.edu; K. D. Zamba: gideon-zamba@uiowa.edu; Matthew Rizzo: matthew-rizzo@uiowa.edu

[a] Department of Biostatistics, University of Iowa, Iowa City, IA, 52242, USA

[b] Department of Neurology, University of Iowa, Iowa City, IA, 52242, USA

[c] Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA, 52242, USA

[d] Public Policy Center, University of Iowa, Iowa City, IA, 52242, USA

Abstract

In driving studies based on simulators and instrumented vehicles, specific models are needed to capture key aspects of driving data such as lateral control. We propose a model that uses weighted polynomial projections to predict each data point from the previous three time points, and accommodates the attempts of the drivers to re-center the vehicle before crossing the borders of the traffic lane. Our model also allows the possibility that average position within the lane may vary from driver to driver. We demonstrate how to fit the model using standard statistical procedures available in software packages such as SAS. We used a fixed-based driving simulator to obtain data from 67 drivers with Alzheimer's disease and 128 elderly drivers without dementia. Using these data, we estimated the subject-specific parameters of our model, and we compared the two groups with respect to these parameters. We found that the parameters based on our model were able to distinguish between the groups in an interpretable manner. Hence, this model may be a useful tool to define outcomes measures for observational and interventional driving studies.

Keywords

Time Series; Natural Bounds; Entropy; Alzheimer's Disease

1. INTRODUCTION

In studies using driving simulators and instrumented vehicles, information on lateral position is measured and recorded at high frequencies (e.g., 10–60 frames/sec), allowing investigators to analyze lateral vehicular control. It is often desirous to reduce the complex time-series patterns of lateral position to individual metrics of lateral control, so that important research hypotheses can be tested in a straightforward manner. For example, regression models and other traditional statistical methods could be used to predict lateral control parameters as a function of such covariates as age, gender, disease status,

Please send correspondence and reprint requests to: Jeffrey Dawson, ScD, Department of Biostatistics, University of Iowa, 200 Hawkins Drive, C22 GH, Iowa City, IA 52242, USA, Tel #: 319-384-5023, Fax #: 319-384-5018, jeffrey-dawson@uiowa.edu.

intervention status, environmental conditions, and measures of vision, cognition, and motor skills.

Many approaches already exist to reduce the high-frequency lateral position data to individual parameters. Some are very simple and straightforward, such as the rate at which the lane markings are crossed and the overall standard deviation of lane position. Others involve complex time-series analyses, which seem to provide a natural foundation for modeling lateral control data. In general, time series procedures can be classified as either frequency domain (e.g., Fourier transform/spectral analyses, wavelet analyses) or time domain (e.g., autoregressive models, moving average models, state space models). In this report, we are primarily interested in modeling the temporal progression of a lateral control series, as opposed to characterizing the prominent periodicities and frequencies of the series; hence, our emphasis will be on the time domain approach.

To investigate and compare the statistical properties of these approaches, and to propose additional approaches for consideration, specific models of lateral position data are needed. Such models should accommodate the fact that lane boundaries exist that encourage the driver to re-center the vehicle as the boundaries are approached or crossed. These models should also allow individual drivers to have their preferred position within the traffic lane. In this paper, we propose a class of time-series models which meet these criteria, and we show how to fit a specific subclass of this model using available statistical software. We then use data from a study of elderly drivers to illustrate how to compare two groups (those with and without Alzheimer's disease) based on estimated parameters from our model, and we contrast those results with those based on other existing methods.

2. METHODS

2.1 Proposed Model

For an individual drive, assume that the lane position at time t is denoted by Y_t. On a coordinate plane, let the time $t = 0, 1, 2,\ldots, T$ be represented on the horizontal axis and Y_t be represented on the vertical axis (see Figures 1 and 2). Let the value of $Y_t = 0$ correspond to situations when the center of the vehicle is in the center of the driving lane, let values of $Y_t > 0$ correspond to when it is left of center (from the driver's perspective), and let values $Y_t < 0$ correspond to when it is right of center. Hence, increasing values of Y indicate that the vehicle is heading towards the left shoulder, while decreasing values of Y indicate that the vehicle is heading towards the right shoulder.

Our general model is a third-order autoregressive time series model (Kendall and Ord, 1990) with a signed error term. That is, for $t>3$, let

$$Y_t=g(Y_{t-1}, Y_{t-2}, Y_{t-3})+|e_t|I_t, \tag{1}$$

$$e_t \sim N(0, \sigma^2), \tag{2}$$

and

$$p_t=prob\,(I_t=-1), esle\ I_t=1. \tag{3}$$

In this general model, *g(.)* is an unknown function which predicts the lateral position at time t based on the position observed at the three previous time points; e_t is normally distributed residuals (errors) between observed and predicted position at time t; σ^2 is the variance of e_t; and I_t is a sign indicator, equaling −1 and 1 with probability p_t and $1-p_t$, respectively. In this general model, the functional form of p_t is unspecified. However, when modeling the lateral position of a safe driver, it would be logical to have p_t increase as Y_t goes from 0 to positive values, since high values of Y_t would indicate that the car is drifting to the left out of the correct traffic lane, so a high probability of a negative error (i.e., $I_t = -1$) would be needed to re-center the vehicle towards the right. Similarly, if Y_t is decreasing from 0 to negative values, this indicates that the driver is drifting towards the right shoulder, so one would want p_t to be low, so that the probability of a positive error would be high, causing the vehicle to re-center back towards the left.

With Equations 1–3 as a general model, we now introduce a series of reparameterizations and constraints to produce a specific subclass of models. We reparameterize the vector, $(Y_{t-1}, Y_{t-2}, Y_{t-3})$, to the vector, (W_{1t}, W_{2t}, W_{3t}), by letting

$$W_{1t}=Y_{t-1}, \tag{4}$$

$$W_{2t}=Y_{t-1}+(Y_{t-1}-Y_{t-3})/2, \tag{5}$$

and

$$W_{3t}=3Y_{t-1}-3Y_{t-2}+Y_{t-3}. \tag{6}$$

In this reparameterization, W_{1t} is a flat projection from time $t-1$ to time t based on the previous value; W_{2t} is a linear projection based on the previous three values (the middle value is ignored in a slope calculation when the points are evenly spaced); and W_{3t} is a quadratic projection based on the previous three values. We also replace the general function *g(.)* with a partially-specified function *f(.)*, which is a linear combination of the above projections, i.e.,

$$g(Y_{t-1},Y_{t-2},Y_{t-3})=f(W_{1t},W_{2t},W_{3t})=\beta_1 W_{1t}+\beta_2 W_{2t}+\beta_3 W_{3t}. \tag{7}$$

It can be shown that the (W_{1t}, W_{2t}, W_{3t}) vector is a linear transformation of the general $(Y_{t-1}, Y_{t-2}, Y_{t-3})$ vector; hence, this reparameterization does not really change the model. However, by considering the overall projection as a linear combination of flat, linear, and quadratic projections, it may add insight to the interpretation of the model. As a hypothetical example, Figure 3 shows four potential predictions of the value at time $t=4$ as a function of the values at times $t=1$, $t=2$, and $t=3$. If one driver's lateral position is well-modeled by the quadratic projection (i.e., $\beta_1=0, \beta_2=0, \beta_3=1$; see the top dashed line in Figure 2), while another's position is well-modeled by a linear projection ($\beta_1=0, \beta_2=1, \beta_3=0$; the second-from-top dashed line in Figure 2), then it could be interpreted that the first subject may have worse driving, since variations in lateral position tend to be exaggerated. If a third driver's data is well-modeled by the projection setting of ($\beta_1=0, \beta_2=1/3, \beta_3=2/3$), as shown in the dotted line in Figure 2, then that driver's driving performance would be judged as being between the other two.

In Equations 4–7, the functional form of *f(.)* is assumed only to be linear in β_1, β_2, and β_3. However, the representation shown in Figure 3 suggests that it may be reasonable to predict future lateral position by using weighted averages of the flat, linear, and quadratic projections. This approach was used in simulations in previous studies (Dawson et al, 2006). In order to force this subclass of linear functions, we impose the following constraints:

$$\beta_1+\beta_2+\beta_3=1 \tag{8}$$

and

$$\beta_1 \geq 0, \beta_2 \geq 0, \text{ and } \beta_3 \geq 0. \tag{9}$$

In addition to the above constraints, we need to specify a functional form of p_t, which is the probability that a negative difference will be seen when subtracting the predicted observation from the observed observation at time *t*. Many functional forms could be used, but we propose using a simple logistic model:

$$log(p_t/[1-p_t])=\gamma_0+\gamma_1 Y_{t-1}, \tag{10}$$

where log(.) is the natural log function (base *e*); γ_0 is the intercept of the logistic model; γ_1 is the slope of the logistic model; and Y_{t-1} is the observed lateral position at the previous time point.

Note that the higher the value of γ_1, the greater the tendency for a driver to turn back towards the center as the vehicle approaches or crosses a lane boundary. Hence, this could be termed a "re-centering" parameter, and high values of γ_1 would tend to indicate better drivers. The intercept, γ_0, can be thought of as a default position parameter, in that it accommodates the manner in which some drivers tend to stay on center, while others tend to ride closer to one lane boundary or the other. When γ_0=0, the vehicle has an average position at the center of the lane, while positive values of γ_0 will correspond to the vehicle tending to be right of center and negative values of γ_0 will correspond to the vehicle tending to be left of center.

Part of the motivation for our proposed model was the fact that drivers feel a greater need to make correction in their steering when distance to the lane boundary decreases. Although our model does not use this distance as a parameter, certain distance measures of interest can be estimated based on the γ parameters of the model. Specifically, additional insight to the driving behavior of an individual can be found solving Equation 10 for Y_{t-1}, i.e.,

$$Y_{t-1}=[log(p_t/[1-p_t])-\gamma_0]/\gamma_1. \tag{11}$$

Hence, after estimates of γ_0 and γ_1 are obtained, one can find the lateral lane position that would correspond to specific probability of having a negative residual. For example, the value of Y_{t-1} when p_t = 0.50 should correspond approximately to the driver's average position in the lane, since the error term has a 50/50 chance of being positive and negative. The value when p_t = 0.05 would indicate the lateral position where there is only a 5% chance of a negative residual (i.e., which would tend to lead the vehicle back towards to the left), and the value when p_t = 0.95 would indicate the lateral position where there is a 95% chance of a negative residual (i.e., tending to turn back to the right). We will label these three lateral positions obtained in this manner, as the "lateral mid-point" and the "5/95 boundaries."

2.2 Fitting the Model

To illustrate a method for estimating the parameters of our model, we now provide a step-by-step guide for fitting our model to an individual's data in SAS (2007, SAS Institute, Cary, NC). However, a variety of software packages could be used. The steps are as follows:

1. Create a dataset with time (t) and lane position (y), sorted by time.
2. Calculate the first-, second-, and third-order lags of y.
3. Based on the lags, create the flat, linear, and quadratic projections using equations 4–6, naming them, say, *flat*, *lin*, and *quad*.
4. To constrain the β estimates to sum to 1, estimate two of the three, so that the third can be found by subtraction. For example, create the following variables, y*_flat = y - flat*, *lin_flat = lin - flat,* and *quad_flat = quad – flat,* and then fit the model, "*y_flat = quad_flat lin_flat*," without an intercept. This regression model does not estimate β_1 directly, but this term will be obtained indirectly by subtraction.
5. Check if the estimates of β_2, β_3, and $\beta_2+\beta_3$ are each in [0,1] range. If all three of these quantities are in this range, then accept the estimates as they are and obtain the estimate of β_1 by subtraction. If they are out of range, then remap them to the parameter space appropriately. For example, if an estimate of β_2 or β_3 is negative, then assign it the value of 0. If $\hat{\beta}_1+\hat{\beta}_2>1$, then subtract $(\hat{\beta}_1+\hat{\beta}_2 - 1)/2$ from each so that they sum to exactly 1.
6. Based on the estimated values of β_1, β_2, and β_3, calculate predicted values of y, that is, $\hat{y} = (flat)\hat{\beta}_1 + (lin)\hat{\beta}_2 + (quad)\hat{\beta}_3$. Calculate residuals, $y - \hat{y}$, noting the sign.
7. Calculate the standard deviation of the residuals, and let this be an estimate of σ.
8. Create an indicator variable for negative residuals (i.e., 1 for negative and 0 for positive), and let that be predicted by the first-order lag of y in a logistic regression model. The intercept and slope terms from the logistic model correspond to estimates of γ_0 and γ_1, respectively.

For several drivers, perform all of the above operations separately for each driver (using "by" statements in SAS), and then look for associations between covariates of interest and the parameters of the model using the appropriate statistical procedure. An example of the SAS syntax for the above steps is given in the Appendix.

2.3 Entropy

Boer (2000) proposed an entropy metric of steering variability to measure the effect of excess workload on drivers. Dawson et al (2006) illustrated that this could also be applied to lane position data. In Boer's method, a second-order Taylor series expansion is used to predict the value of the response at time t, based on the three immediately preceding values. It can be shown that this Taylor series prediction is equivalent to Equations 4–7 in the present paper if $\beta_1=0$, $\beta_2=1/3$, and $\beta_3=2/3$ (see Figure 3). From this projection, a set of prediction errors (residuals) are calculated. The 95th percentile of the distribution of the predictions errors is labeled as x (Boer used α). Nine bins are then defined according to eight borders: $-5x$, $-2.5x$, $-x$, $-0.5x$, $0.5x$, x, $2.5x$, and $5x$. Finally, these bins are used to calculate entropy, as $H = \Sigma -P_i \log_9 (P_i)$, where H is the entropy, scaled between 0 and 1, and P_i is the observed proportion of prediction errors in bin i (not the same as p_t). The bins are often defined from a baseline situation, with the entropy calculated for a second interval; however, the method can also be applied to a single interval of data (Boer, personal communications). For comparing two groups, an additional modification of this entropy method is to take the average 95th percentile of prediction error distributions within one group, and use this *x* to define fixed bin widths that are used for calculating the entropy measure.

2.4 Data Example

We obtained data from a driving simulator known as SIREN ("Simulator for Interdisciplinary Research in Ergonomics and Neuroscience"). This fixed-base simulator captures lane position data at 30 frames per second (Rizzo, 2004). We enrolled 67 drivers with Alzheimer's disease (AD) and 128 elderly drivers who were neurologically normal. See Uc et al (2006) for more details of the study. All participants provided signed informed consent and the study was approved by the University of Iowa Institutional Review Board.

We used data from a straight-road segment of approximately 3.2 kilometers in length. Figure 1 shows 60 seconds of data from an AD subject, illustrating excessive variability and several lane crossings. Figure 2 shows data from an elder driver without AD, with less variability. Note that since the vertical axis is in meters and the horizontal axis is in seconds, the graphs are not truly geographical, thus giving an exaggerated picture of the lateral swerves.

Our outcome of interest was the lateral position of the center of the vehicle. Our original data was sampled at 30 Hz. For the purposes of our analyses, we averaged the lane position data over consecutive, non-overlapping five-frame blocks, resulting in 6-Hz sampling. We fit our model to each driver using the algorithm presented in Section 2.2, estimating his/her values of the regression coefficients for projections (β_1, β_2, and β_3), the error variance (σ^2), the default position parameter (γ_0), and the re-centering vigilance parameter (γ_1). We also estimated Boer's entropy for each subject, both using subject-specific bins, as well as a fixed-bin approach based on the average 95th percentile of the prediction error distributions in the non-AD group. We also computed the mean, standard deviation, skewness, and kurtosis estimates for each driver's prediction error distribution that was obtained during the entropy calculations. We also computed the simple standard deviation of the lane position, as well as rate at which the wheels of the vehicle crossed the lane markings. Finally, using Equation 11 in the manner described in Section 2.1, we calculated the lateral mid-points and the 5/95 boundaries at the mean values of the estimates of γ_0 and γ_1 within each group.

Each of the measures of variability was calculated for each subject, with the mean and standard deviation of these measures calculated for each group. The groups were compared using Wilcoxon Rank-sum tests (a.k.a. "Mann-Whitney tests), so that the comparisons would be robust to the influence of outliers.

3. RESULTS

Table 1 displays the results of our comparisons. We found that those with AD had, on average, higher quadratic components and lower linear components than the non-AD controls, indicating exaggerated lateral movements of the vehicle ($p<0.0001$). They also had lower estimates of γ_1, indicating less vigilance in re-centering the vehicle, ($p<0.0001$). The average estimates of γ_0 were positive in both groups, indicating a tendency to be closer to the right lane marker than to the left lane marker, which is generally a safe strategy when there is oncoming traffic in the left lane and no obstacles in the shoulder, as was the case in our simulated scenario. Based on the estimated values of γ_0, the non-AD controls were slightly more to the right than the AD subjects ($p=0.045$).

The model-based residual standard deviation, the entropy measures, and the descriptive statistics based on the prediction errors from the entropy calculations all showed no significant differences between the two groups ($p>0.05$ in all cases). The standard deviation of lane position and the lane crossing rate were both higher in the Alzheimer's disease group ($p=0.0004$ and 0.037, respectively), but the effect sizes of these differences was smaller than three of the four significant parameter differences found by our model.

We found that the typical lateral mid-point was −0.26 meters for AD drivers and −0.28 meters for non-AD drivers, showing the tendency for both sets of drivers to be slightly right of center, with the non-AD slightly more offset than the AD drivers. This finding is consistent with the interpretation of γ_0 given above. We also found that the typical 5/95 boundaries are −2.06/1.54 meters for the AD drivers and −1.56/1.01 meters for the non-AD drivers. Hence, the AD drivers tend to be approximately 0.5 meters further away from the center of the lane before they re-center with the same level of certainty (95%) as non-AD drivers.

4. DISCUSSION

Our proposed model was able to distinguish between AD and non-AD subjects with respect to lateral control. The parameters that differed between the groups lend themselves to meaningful clinical interpretations. Specifically, since those with AD had higher quadratic components and lower linear components, this suggested that AD drivers maintained less control of the vehicle, in that their lateral movement trends were exaggerated and less less-dampened compared to non-AD drivers. AD drivers were also less vigilant in moving the vehicle back towards the center of the lane when the lane boundaries were approached or crossed, indicating a greater potential of lane crossing. Finally, AD drivers were slightly closer to oncoming traffic in the left lane, which could raise the risk of head-on collisions.

Our model does not explicitly represent human perceptual motor control processes, but appears to be adequate approximation to observed lateral lane position. However, compared to simple analyses, such as the standard deviation of lane position and the lane crossing rate, the model does give separate components that can be interpreted so that driver behavior can be better understood. The three general components of our model could be thought of as the projection or filter component (captured by g(.)), the noise magnitude component (indexed by σ^2), and the correction component (p_t). When comparing AD versus non-AD drivers, we found no difference in the noise component, but notable differences in the projection component (more exaggerated in AD) and in the correction component (less vigilance). The differences in both the projection and correction components may reflect visuomotor impairments in AD, and there may also be visuoperception impairments that relate to the correction (re-centering) parameters.

The model-based residual standard deviation (our estimate of σ) is related conceptually to the standard deviation of the prediction error distribution that is part of the entropy calculation. In our model, we let the data determine, for each subject, the best-fitting linear combination of the flat, linear, and quadratic projections to use for predicting each data point. In the entropy calculation, the coefficients are set *a priori* to be $\beta_1=0$, $\beta_2=1/3$, and $\beta_3=2/3$ for every subject. Hence, it is not surprising that the standard deviations of the residuals from our model were much smaller than those based on the entropy measure (0.0046 vs. 0.0145 meters). No between-group differences were found when comparing the entropy-based prediction error distributions in numerous ways (subject-specific entropy, fixed-bin entropy, and the mean, standard deviation, skewness, and kurtosis of the prediction errors).

The parameter estimates in our model can be greatly influenced by the sampling rate of the time-series data. In our data example, we reduced our 30 Hz data frames down to 6 Hz data blocks of averages before fitting our model. We did this for two reasons. First, the resulting 167-msec block width is on the same order of magnitude as the minimum time of 150 msec reported to be necessary for humans to make corrections when doing visuomotor tracking (Jagacinski and Hah, 1988). Second, preliminary analysis revealed that with such a high original sampling rate, and with finite precision in our measurement scale of lateral position (1 cm), it was very common to have several points in a row with an identical value, followed by a step up or down to the next possible value. With the raw data following this step function, the projection based on three previous points was often exactly what was observed, resulting in many “0” values for the residual errors in our model, as well as for the prediction errors based on the entropy approach. Such discreteness was a direct violation of our model, which assumed normal errors. However, when using averages over five frames, the resulting data allowed a continuum of residual errors.

A number of methodological issues remain to be investigated. For example, even though our model estimates gave interpretable results when comparing AD with non-AD drivers, we have not yet formally investigated the statistical properties of our multi-stage method of fitting our model. Thorough simulations are needed to examine the coverage performance of confidence intervals and hypothesis tests (Type I and II error rates) based on the algorithm we used to fit our model. Also, the performance of our model should be compared with the performance of many other time-series models and approaches (e.g., see Kedem and Fokianos, 2002; Donges, 1978). Finally, although we found our model parameters to give better discrimination between the groups than Boer’s entropy, it is not clear whether this would be the case when looking at high-workload versus low-workload situations, as described by Boer (2004). There are also other versions of “entropy” that could be considered as candidates (Richman and Moorman, 2000).

In this paper, we proposed a general model (Equations 1–3), as well as specific reparameterizations (4–7), constraints (8–9), and a functional form of the re-centering process (10). Other specific subclasses of the model could be proposed. In particular, it may be that the two-parameter logistic model for re-centering (Equation 10) is overly restrictive, and a more general functional form may be able to accommodate the possibility of drivers recentering faster as they drift left versus when they drift right, due to the worse potential consequences of the former (e.g., head-on collisions).

In our data example, some of the parameters of our model had greater effect sizes than the rate of vehicle crossings. This suggests the possibility that the application of our model to short time intervals may be able to predict lane crossings in future intervals. Our proposed model has the potential to help assess driver safety, to provide inputs in vehicle safety systems, and to provide the basis of outcome measures in interventional studies.

Acknowledgments

The authors would like to thank the entire research team of the Division of Neuroergonomics at the University of Iowa. This work was supported by NIA awards AG17177 and AG15071.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Appendix

```
* Begin with a dataset called 'drivedata', with the following variables:
  ID: Driver identification number
  Y:  The lateral position
  Time:     For each driver, this must take on values 1, 2, 3, etc.;

data d1; Set drivedata;

proc sort; by id time;                                  * STEP 1;

data d2;                                                * STEPS 2 & 3;
 set d1;
  ylag1=lag1(y);
  ylag2=lag2(y);
  ylag3=lag3(y);
* The next line is to ensure that the final 3 frames of data from one
      driver do not contaminate the next driver;
  if time<4 then do;  ylag1=.; ylag2=.; ylag3=.; end;
 flat=ylag1;
 lin=ylag1+(ylag1-ylag3)/2;
 quad=ylag3-3*ylag2+3*ylag1;
 y_flat=y-flat;
 lin_flat=linear-flat;
 quad_flat=quad-flat;

proc sort; by id time;

proc reg  data=d2 outest=d3 covout noprint;             * STEP 4;
model y_flat = quad_flat lin_flat  / noint;
by id;

data d4;                                                * STEP 5;
 set d3;
 a=quad_flat; b=lin_flat; aold=a; bold=b;
 if a<0 and b<0 then do; a=0; b=0; end;
 if b<0 and a<1 and a>0 then do; b=0; end;
 if a<0 and b<1 and b>0 then do; a=0; end;
 if a>1 and b<a-1 then do; a=1; b=0; end;
 if b>1 and b>a+1 then do; a=0; b=1; end;
 if a>0 and b>0 and a+b>1 then do;
    a=aold-.5*((aold+bold)-1); b=bold-.5*((aold+bold)-1); end;
 c=1-a-b;
 beta3=a; beta2=b; beta1=c;
 if _type_ = "PARMS";

proc sort; by id;

data d5; merge d2 d4; by id;

data d6; set d5;                                        * STEP 6;
resid=y-(beta1*flat+beta2*lin+beta3*quad);

proc sort; by id beta1 beta2 beta3;

proc means data=d6 noprint;                             * STEP 7;
```

```
var resid;
 output out=d7 std=sdresid;
 by id beta1 beta2 beta3;

data d8; set d7; drop _type_;

data d9;                                              * STEP 8;
 set d6;
 if resid<0 then negerr=1;
 if resid>0 then negerr=0;

proc logistic descending data=d9 outest=d10 covout noprint;
model negerr=flat;
by id;

data d11;
 set d10;
 gamma0=Intercept;
 gamma1=flat;
 if _type_='PARMS';
 keep id gamma0 gamma1;

proc sort; by id;

data finaldataset;
 merge d8 d11;
  by id; run;
```

References

Boer, ER. Behavioral entropy as an index of workload. 44th Annual Meeting of the Human Factors and Ergonomics Society (HFES2000); San Diego, CA. 2000.

Dawson, JD.; Cavanaugh, JE.; Zamba, KD.; Rizzo, M. Measuring lateral control in driving studies. Biometric Society (ENAR) Annual Meeting; Tampa, FL. 2006.

Donges E. A two-level model of driver steering behavior. Human Factors 1978;20(6):691–707.

Jagacinski RJ, Hah S. Progression-regression effects in tracking repeated patterns. Journal of Experimental Psychology 1988;14(1):77–88. [PubMed: 2964507]

Kendall, MG.; Ord, JK. Time Series. 3. Edward Arnold; London: 1990.

Kedem, B.; Fokianos, K. Regression Models for Time Series Analysis. John Wiley; New York: 2002.

Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology – Heart and Circulatory Physiology 2000;278:H2039–H2049. [PubMed: 10843903]

Rizzo, M. Safe and unsafe driving. In: Rizzo, M.; Eslinger, PJ., editors. Principles and Practice of Behavioral Neurology and Neuropsychology. WB Saunders; Philadelphia, PA: 2004. p. 197-222.

Uc EY, Rizzo M, Anderson SW, Shi Q, Dawson JD. Unsafe rear-end collision avoidance in Alzheimer's disease. Journal of the Neurological Sciences 2006;251:35–43. [PubMed: 17049360]

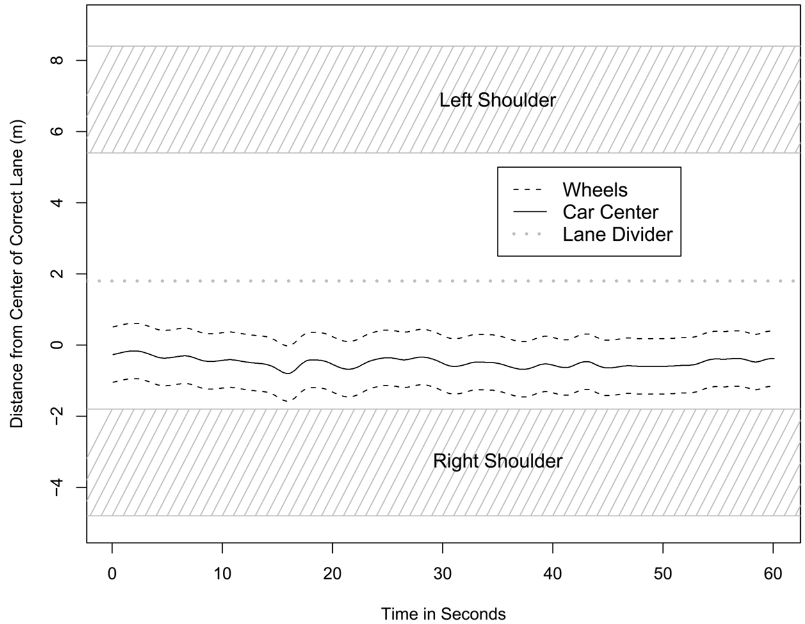

Figure 1.
Example of lane position data for a driver with Alzheimer's disease.

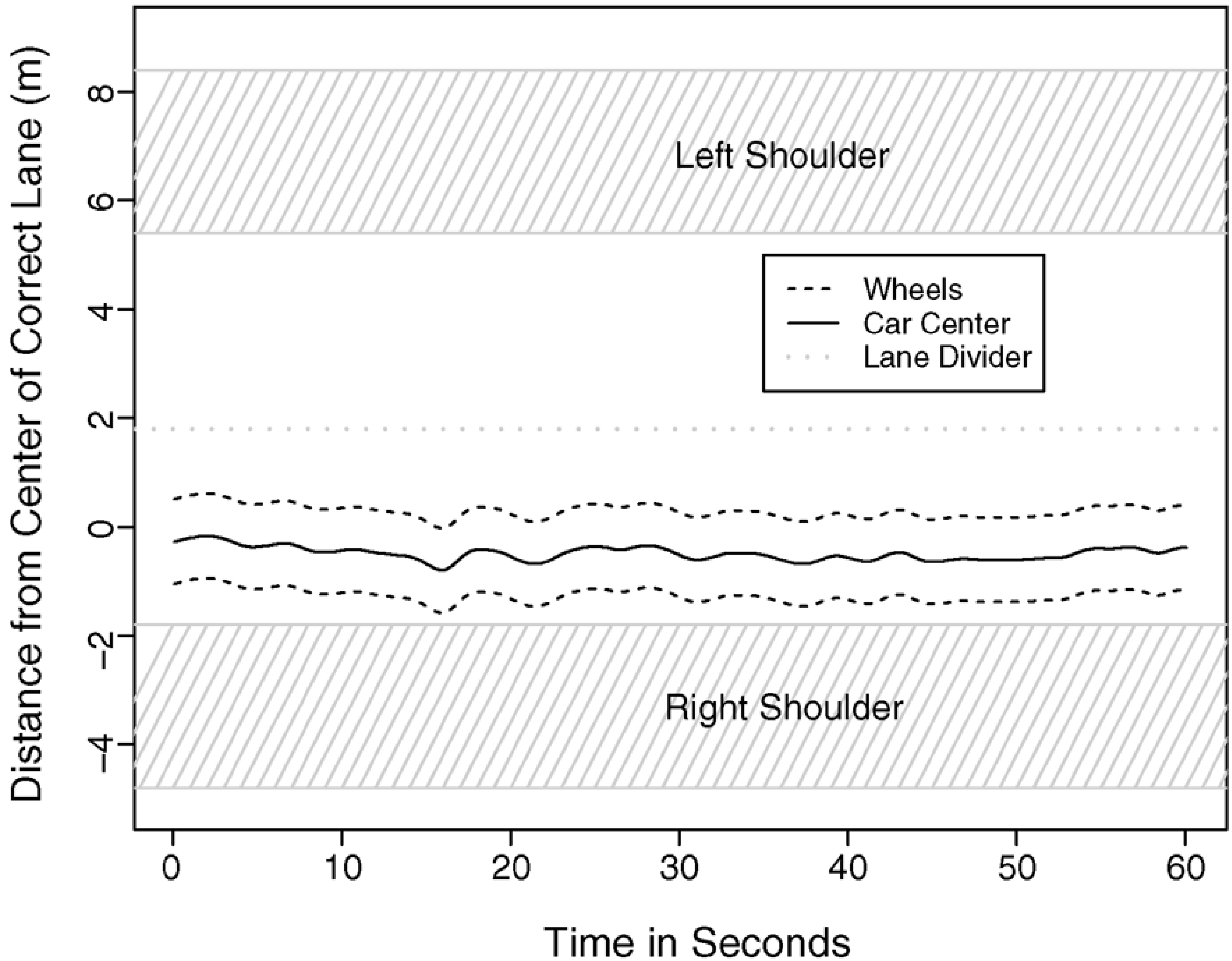

Figure 2.
Example of lane position data for an elderly driver without Alzheimer's disease.

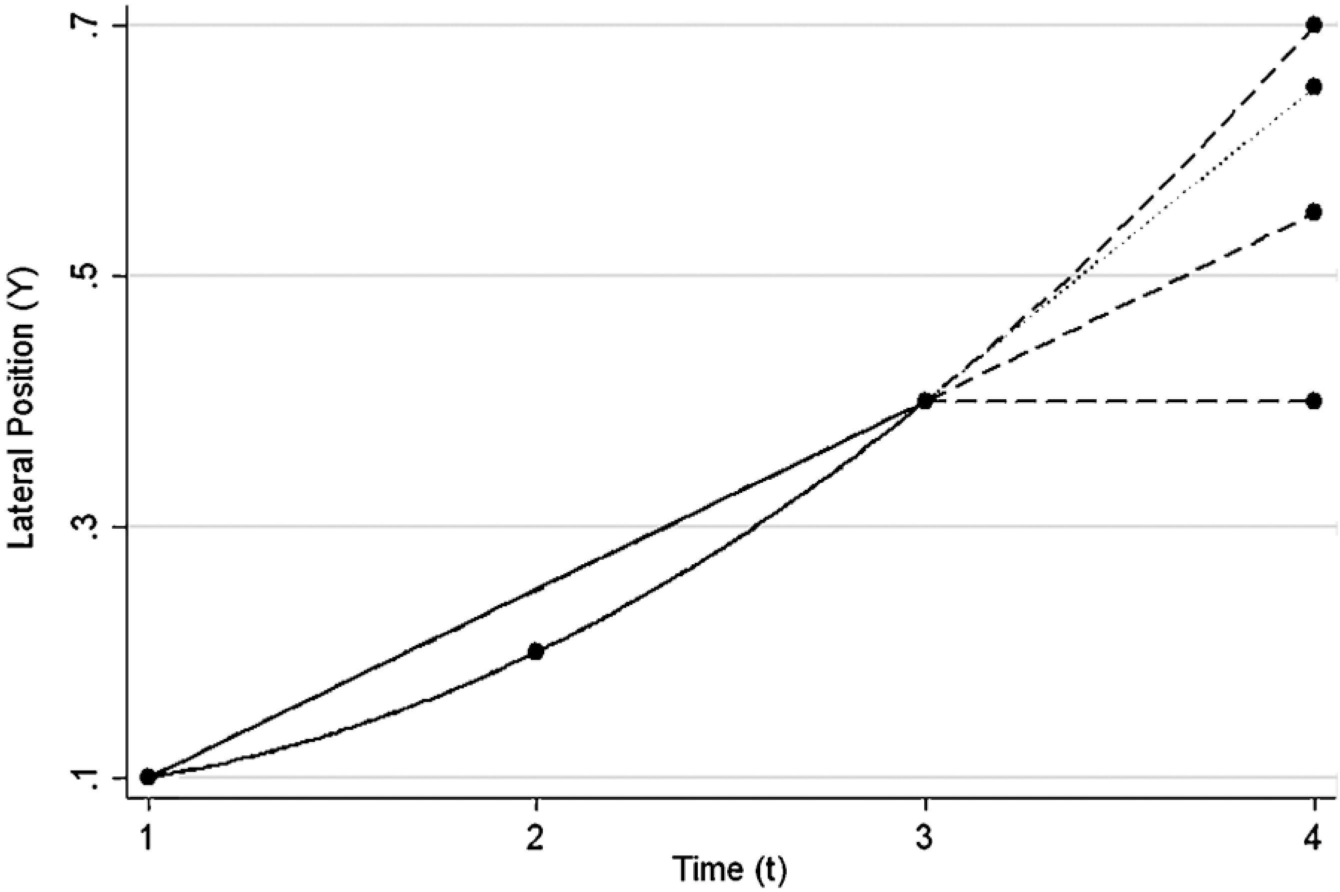

Figure 3.

Hypothetical examples of projections based on three previous time points. From bottom to top, the dashed lines show the flat, linear, and quadratic projections, corresponding to coefficient sets of ($\beta_1=1, \beta_2=0, \beta_3=1$), ($\beta_1=0, \beta_2=1, \beta_3=0$), and ($\beta_1=0, \beta_2=0, \beta_3=1$), respectively. The dotted line shows a weighted average of linear and quadratic projections, ($\beta_1=0, \beta_2=1/3, \beta_3=2/3$).

Table 1

Results of Between-Group Comparisons

Parameter Being Estimated, or Variability Measure	Mean (SD), Median		Wilcoxon Rank-Sum Statistics (AD vs. Non-AD)
	AD Subjects (n=67)	Non-AD Controls (n=128)	
β_1 (Flat Component)	0.052 (.020)	.055 (.018)	−1.03
β_2 (Linear Component)	0.31 (0.24)	0.47 (0.25)	−4.09***
β_3 (Quad. Component)	0.64 (0.24)	0.48 (0.25)	4.19***
σ	0.0046 (0.0013)	0.0046 (0.0008)	0.25
γ_0 (Default Position)	0.42 (0.54)	0.63 (0.79)	−2.05*
γ_1 (Re-centering)	1.63 (1.14)	2.29 (1.35)	−3.78**
Subj.-Specific Entropy	0.59 (0.03)	0.60 (0.03)	−1.78
Fixed-Bin Entropy	0.52 (0.16)	0.56 (0.08)	−0.11
Mean Pred. Error (PE)	0.00033 (0.00063)	0.00033 (0.00056)	−0.29
SD of PE	0.0150 (0.0068)	0.0145 (0.0056)	−0.12
Skewness of PE	4.15 (7.85)	3.94 (6.12)	−0.19
Kurtosis of PE	102.0 (109.7)	77.4 (85.7)	0.97
SD of Lane Position	0.29 (0.10)	0.25 (0.16)	3.53**
Lane Crossings per Minute	0.85 (1.71)	0.39 (0.96)	2.08*

*p<0.05;

**p<0.001;

***p<0.0001