DEVELOPMENT OF A RIDER SIZE AND POSITION MODEL TO DETERMINE MOTORCYCLE PROTECTIVE DEVICE TEST CONDITIONS

R. Michael Van Auken
Scott A. Kebschull
Peter C. Broen
John W. Zellner
Dynamic Research, Inc.
United States
Nicholas M. Rogers
International Motorcycle Manufacturers Association
Switzerland
Paper Number 05-0392

ABSTRACT

The benefits and risks of potential motorcycle protective devices (such as airbags) may depend on the pre-crash position of the rider on the motorcycle. Therefore an understanding of the range of riding positions is needed for research into the risks and benefits of these devices. A statistical model was developed that describes the range motorcycle riding positions, in terms of mean, variance, and correlation parameters; as a function of rider stature, motorcycle-rider interface geometry (seat, hand grips, footrest), and geographic region, based on data collected from Japan, Europe (the Netherlands), and the United States. The rider position and motorcycle-rider interface geometry data was digitized from images of 1390 riders as they were riding on public roadways. A graphical user interface was developed to enable a user to select from and view the range of riding positions described by the model.

INTRODUCTION

Background

The benefits and risks of potential motorcycle protective devices (such as airbags) may depend on the pre-crash position of the rider on the motorcycle (See Rogers and Zellner [1]). Therefore an understanding of the range of riding positions is needed for research into the risks and benefits of these devices.

In addition, motorcycle seating and control layout may affect other, various vehicle attributes, such as comfort, aerodynamics, visibility, stability and so on, and therefore the interaction with ride size and position are of general interest. For example, Reed et al. [2] have developed a similar model for automobile driving posture.

Objectives

The objectives of this study were to determine the range of motorcycle riding positions, as a function of rider stature, motorcycle-rider interface geometry (seat, hand grips, footrest), and possibly geographic region.

Approach

The distribution of motorcycle riding positions was quantified in terms of a mean and variation of the rider back angle and hip position in terms of the rider the hand grip location and seat height relative to the footrests, and rider statue. The mean location was determined by a regression analysis, assuming a constant (homogeneous) variation in riding position about the mean. The homogeneity of the variation were then assessed by subdividing the data into quartiles.

RIDER POSITION DATABASE

A rider position database was prepared comprising 1390 observations of riders operating motorcycles on public roads in Japan, the Netherlands, and the United States (Ohio) as indicated in Table 1. There were 56 variables for each observation, comprising coordinates of points on the rider and motorcycle that were digitized from video still images, and derived measures such as rider stature. Of the 1390 cases, 139 cases were not used in the analysis because they were outliers (> 5 standard deviations) or had missing data (e.g., a point was not visible on the still image and therefore could not be digitized).

Data Sources

The video and digitized data were provided by Honda, Kawasaki, Suzuki, and Yamaha of Japan, Harley-Davidson (USA), and Yamaha (Europe). The sources and regions are summarized in Table 1.

Table 1.
Rider Position Database Sources and Regions

Source	Region	Number of Observations
Harley-Davidson	US (Ohio)	218
Honda	Japan	150
Kawasaki	Japan	150
Suzuki	Japan	159
Yamaha (Japan)	Japan	154
Yamaha Motor Europe	Europe (Netherlands)	559
Total		1390

Rider Position Variables

The variables of interest for this analysis are as follows:

- Independent variables:
  - θ_{back} is the rider back angle in degrees relative to the vertical. A positive value indicates that the rider is leaning forward.
  - x_{hip} is the rider longitudinal hip position relative to the footrest.
- Dependent variables:
  - z_{seat} is the seat height (m),
  - x_{grip} is the rider longitudinal hand grip position (m),
  - z_{grip} is the rider vertical hand grip position (m),
  - S' is the estimated rider stature (m).

The X and Z coordinates are expressed in meters relative to the location of the motorcycle footrest. The Z-axis is perpendicular to the ground plane and the positive direction is pointed towards the ground. The X-axis is in the plane of symmetry of the motorcycle and perpendicular to the Z-axis; the positive direction is pointed towards the front of the motorcycle.

The rider stature was estimated from the distances between the head center, shoulder point, hip point, knee, and ankle, with corrections for the head height and ankle height.

Distribution of Rider Position Variables

The distributions of the position variables are illustrated in Figures 1 to 4. Figures 1 and 2 are histograms illustrating the univariate distributions for each of the dependent and independent variables. Figures 3 and 4 are scatter plots illustrating the distribution of the rider hip position and back angle vs the independent variables. These figures indicate a wide range of rider positions for which more detailed model is sought.

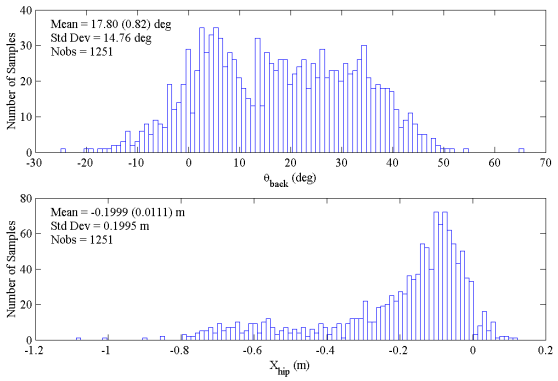

Figure 1. Distribution of rider longitudinal hip position and back angle.

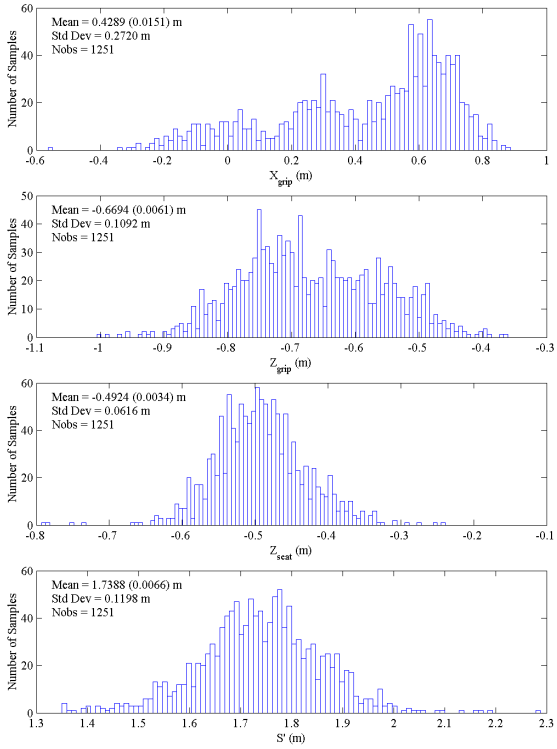

Figure 2. Distribution motorcycle seat height, hand grip location, and rider stature.

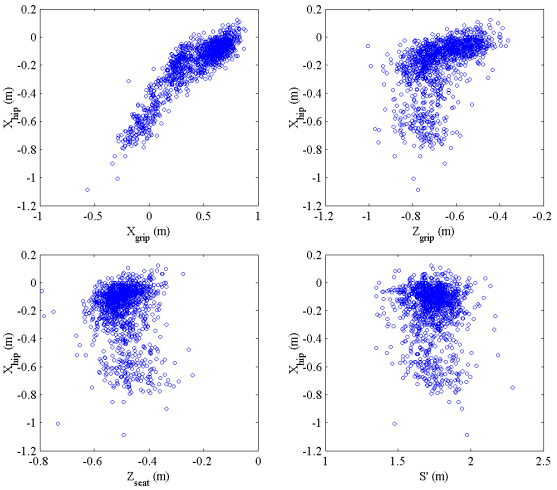

Figure 3. Distribution of rider hip position vs motorcycle seat height, hand grip location, and rider stature.

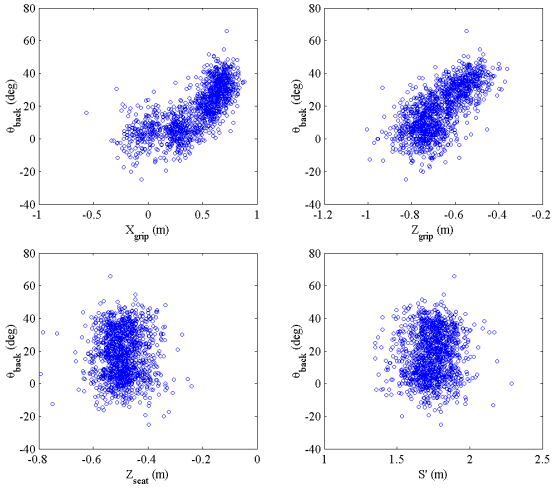

Figure 4. Distribution of rider back angle vs motorcycle seat height, hand grip location, and rider stature.

METHODOLOGY

Assumed Rider Position Distribution

It was assumed that the rider position can be characterized in terms of the longitudinal hip position (x_{hip}) and back angle (θ_{back}) that depend on the motorcycle seat height (z_{seat}) and hand grip position (x_{grip} , z_{grip}), rider stature (S'), region (R), plus some amount of random variation. More specifically, it is assumed that x_{hip} and θ_{back} are normally distributed with mean and variance as follows:

Mean:

$$\begin{aligned} E(x_{hip}) &= \mu_{x_{hip}} = \mathrm{F}_x(x_{grip}, z_{grip}, z_{seat}, S', R) \\ E(\theta_{back}) &= \mu_{\theta_{back}} = \mathrm{F}_\theta(x_{grip}, z_{grip}, z_{seat}, S', R) \end{aligned} \quad (1).$$

Variation:

$$\begin{aligned} E\left(\left(x_{hip} - \mu_{x_{hip}}\right)^2\right) &= \sigma^2_{x_{hip}} \\ E\left(\left(\theta_{back} - \mu_{\theta_{back}}\right)^2\right) &= \sigma^2_{\theta_{back}} \\ E\left(\left(x_{hip} - \mu_{x_{hip}}\right)\left(\theta_{back} - \mu_{\theta_{back}}\right)\right) &= \rho\sigma_{x_{hip}}\sigma_{\theta_{back}} \end{aligned} \quad (2).$$

where

$\mathrm{F}(x)$ is a function of x,

$E(x)$ is the expected value of x,

and where $\sigma_{x_{hip}}$, $\sigma_{\theta_{back}}$, and ρ are constant values to be estimated.

It is furthermore assumed that $\mu_{x_{hip}}$ and $\mu_{\theta_{back}}$ in equation (1) are linear combinations of x_{grip}^{p} , z_{grip}^{p} , z_{seat}^{p} , and $(S')^{p}$, for p=1 and 2, for each region R separately and for all regions as a group. This can be expressed as follows:

$$\begin{aligned}\mu_{x_{hip}} &= \alpha_{0,R} + \alpha_{1,R} x_{grip} + \alpha_{2,R} x_{grip}^2 \\ &\quad + \alpha_{3,R} z_{grip} + \alpha_{4,R} z_{grip}^2 \\ &\quad + \alpha_{5,R} z_{seat} + \alpha_{6,R} z_{seat}^2 \\ &\quad + \alpha_{7,R} S' + \alpha_{8,R} (S')^2 \\ \mu_{\theta_{back}} &= \beta_{0,R} + \beta_{1,R} x_{grip} + \beta_{2,R} x_{grip}^2 \\ &\quad + \beta_{3,R} z_{grip} + \beta_{4,R} z_{grip}^2 \\ &\quad + \beta_{5,R} z_{seat} + \beta_{6,R} z_{seat}^2 \\ &\quad + \beta_{7,R} S' + \beta_{8,R} (S')^2\end{aligned} \quad (3).$$

where the values for $\alpha_{k,R}$ and $\beta_{k,R}$ can be estimated by multivariable linear regression (Draper and Smith [1]).

Table 2.
Estimated rider distribution model coefficients

Distribution Model	Parameter	Japan	Europe	US	All Regions
Mean hip position $\bar{x}_{hip} = a_0 + a_1 x_{grip} + a_2 x_{grip}^2 + a_3 z_{grip} + a_4 z_{grip}^2 + a_5 z_{seat} + a_6 z_{seat}^2 + a_7 S' + a_8 (S')^2$	a_0	-0.42 (0.24)	-0.15 (0.15)	0.26 (0.32)	-0.089 (0.057)
	a_1	1.104 (0.053)	0.931 (0.056)	0.873 (0.071)	1.055 (0.038)
	a_2	-0.667 (0.097)	-0.279 (0.082)	-0.38 (0.13)	-0.599 (0.059)
	a_3	-0.62 (0.66)	0.60 (0.37)	1.08 (0.79)	0
	a_4	-0.41 (0.49)	0.41 (0.28)	0.73 (0.59)	-0.032 (0.041)
	a_5	0	0	0	0
	a_6	0	0	0	-0.103 (0.075)
	a_7	-0.161 (0.053)	-0.116 (0.049)	-0.22 (0.10)	-0.212 (0.037)
	a_8	0	0	0	0
Regression model RPRED statistic		*0.87*	*0.92*	*0.93*	*0.88*
Mean back angle $\bar{\theta}_{back} = b_0 + b_1 x_{grip} + b_2 x_{grip}^2 + b_3 z_{grip} + b_4 z_{grip}^2 + b_5 z_{seat} + b_6 z_{seat}^2 + b_7 S' + b_8 (S')^2$	b_0	105 (26)	13.4 (5.5)	27 (15)	51 (15)
	b_1	-10.2 (6.3)	15.2 (8.4)	17 (11)	6.9 (4.5)
	b_2	57 (11)	9 (11)	17 (18)	30.0 (6.7)
	b_3	282 (76)	0	64 (18)	117 (43)
	b_4	169 (56)	-48.5 (6.0)	0	47 (32)
	b_5	-31 (15)	-41 (12)	-34 (26)	-26.8 (8.7)
	b_6	0	0	0	0
	b_7	0	0	0	0
	b_8	0	0	0	0
Regression model RPRED statistic		*0.63*	*0.72*	*0.75*	*0.68*
Hip position standard deviation	$s_{x_{hip}}$	0.0686	0.0512	0.0585	0.0688
Back angle standard deviation	$s_{\theta_{back}}$	7.93	7.03	8.42	8.28
Correlation	r	-0.554	-0.434	-0.390	-0.432
Number of observations		*563*	*501*	*185*	*1251*

Note 95% confidence intervals are in parenthesis ().

Mean Rider Position Model

Given $a_{k,R}$ and $b_{k,R}$ are unbiased linear regression estimates of $\alpha_{k,R}$ and $\beta_{k,R}$, it follows that

$$\begin{aligned}\bar{x}_{hip} &= a_{0,R} + a_{1,R}x_{grip} + a_{2,R}x_{grip}^2 \\ &\quad + a_{3,R}z_{grip} + a_{4,R}z_{grip}^2 \\ &\quad + a_{5,R}z_{seat} + a_{6,R}z_{seat}^2 \\ &\quad + a_{7,R}S' + a_{8,R}(S')^2 \\ \bar{\theta}_{back} &= b_{0,R} + b_{1,R}x_{grip} + b_{2,R}x_{grip}^2 \\ &\quad + b_{3,R}z_{grip} + b_{4,R}z_{grip}^2 \\ &\quad + b_{5,R}z_{seat} + b_{6,R}z_{seat}^2 \\ &\quad + b_{7,R}S' + b_{8,R}(S')^2\end{aligned} \tag{4}$$

are unbiased estimates of $\mu_{x_{hip}}$ and $\mu_{\theta_{back}}$.

The form of equation (4), with linear and quadratic terms, was chosen in order to model possible nonlinear trends in the data. However, not all of these trends may be present in the data and therefore it is appropriate to remove terms that do not contribute to the "fit and predictive capability" of the model. This was accomplished by evaluating the RPRED statistic for all 255 possible models with different combinations of the input terms. The RPRED statistic is described in Appendix D. The model with the maximum RPRED was then chosen and the coefficients for the terms that were removed were set to 0.

Rider Position Variation Model

The difference between the observed and the mean rider position values can be expressed as

$$\begin{aligned}\Delta x_{hip_k} &= x_{hip_k} - \bar{x}_{hip} \\ \Delta\theta_{back_k} &= \theta_{back_k} - \bar{\theta}_{back}\end{aligned} \tag{5}$$

It then follows that

$$\begin{aligned}s_{x_{hip}}^2 &= \frac{1}{n-5}\sum_{k=1}^{n}\Delta x_{hip_k}^2 \\ s_{\theta_{back}}^2 &= \frac{1}{n-5}\sum_{k=1}^{n}\Delta\theta_{back_k}^2 \\ r &= \frac{1}{(n-5)s_{x_{hip}}s_{\theta_{back}}}\sum_{k=1}^{n}\Delta x_{hip_k}\Delta\theta_{back_k}\end{aligned} \tag{6}$$

are unbiased estimates of $\sigma_{x_{hip}}^2$, $\sigma_{\theta_{back}}^2$, and ρ respectively, where n is the number of observations that were used in the linear regression.

RESULTS FOR ALL GEOGRAPHIC REGIONS COMBINED

The distribution of rider position for all 1251 observations from Japan, Europe, and the US was assessed and the resulting distribution model coefficients are listed in the last column of Table 2. The actual distribution of the data was then compared to the distribution model to verify the assumptions.

Verification of the Assumed Equation for the Mean Rider Position

Scatter plots illustrating the distribution of rider longitudinal hip position (x_{hip}) and back angle (θ_{back}) vs the seat height, hand grip location, and stature, while controlling for the other independent variables, are illustrated in Figure 5 and Figure 6. These plots help to verify the assumed quadratic relationship between the dependent and independent variables that was assumed by equation (3). However, there are some small but statistically significant higher order mean deviations that are observable in Figures 7 and 8.

Verification of the Homogeneous Normal Distribution Assumption

The variations in the dependent rider position variables versus the mean values are illustrated in Figures 7 and 8. The colors of the points in the scatter plot at the bottom of each figure illustrate how the data were divided into four equally sized subsets or quartiles. Histograms of each quartile and the entire data set are illustrated at the top of each figure.

These results in Figure 7 suggest that may be some lack of homogeneity in the Δx_{hip} variation, which is larger in the first quartile and smaller in the fourth quartile.

The back angle variation in Figure 8 appears to be consistent with the assumptions.

Verification of the Homogeneous Correlation Assumption

The correlation between Δx_{hip} and $\Delta\theta_{back}$ is illustrated in Figure 9. The size of the ellipse

represents the 95% confidence interval. The ellipsoid appears to be representative of the distribution.

Figure 10 illustrates the same correlation by quartile, in order to observe the homogeneity of the correlation. The quartiles were determined by $\bar{x}_{hip}$ and $\bar{\theta}_{back}$ as illustrated by the scatter plot at the bottom of Figure 10. The correlation appears to be relatively homogeneous, except for the non-homogeneous variation in Δx_{hip} previously noted for Figure 7.

Comparison of the Modeled and Observed Rider Position Distributions

Figure 11 illustrates the agreement between the modeled and observed rider position distributions. The modeled distribution was calculated from

- the values for a, s, and r, listed in Table 2 that describe the mean, variance, and correlation of the dependent rider position variables as a function of the independent variables;
- the observed distribution of the independent variables (e.g., Figure 2); and
- the assumption that the distribution is normally distributed.

The results in Figure 11 indicate that the modeled distribution is in good agreement with the overall distribution. The results also indicate that there are some higher order variations in the distribution that are not modeled, and this may be attributed to the non-homogenous Δx_{hip} previously indicated, and differences due to geographic region (which are not included in this model).

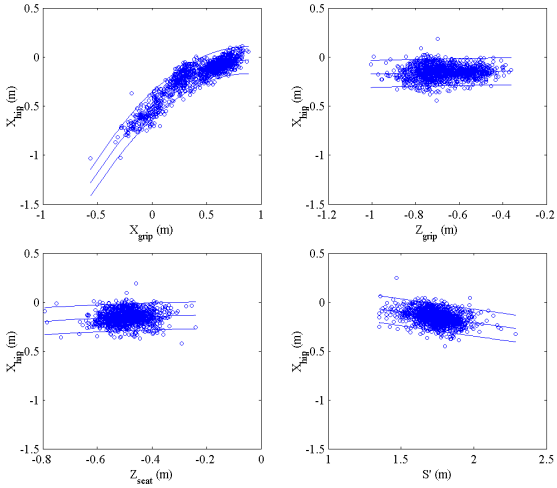

Figure 5. Rider longitudinal hip position vs motorcycle seat height, hand grip location, and rider stature, while controlling for the other independent variables.

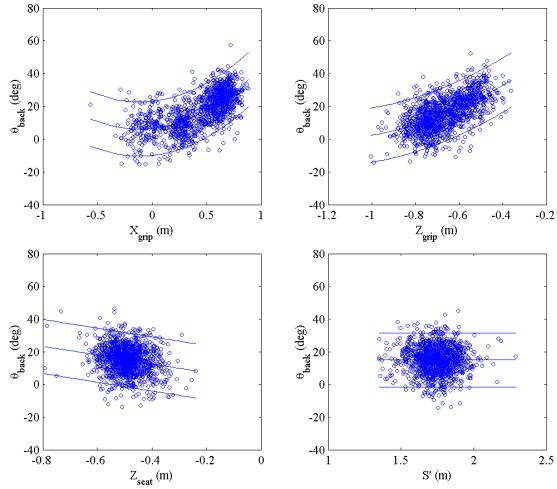

Figure 6. Rider back angle vs motorcycle seat height, hand grip location, and rider stature, while controlling for the other independent variables.

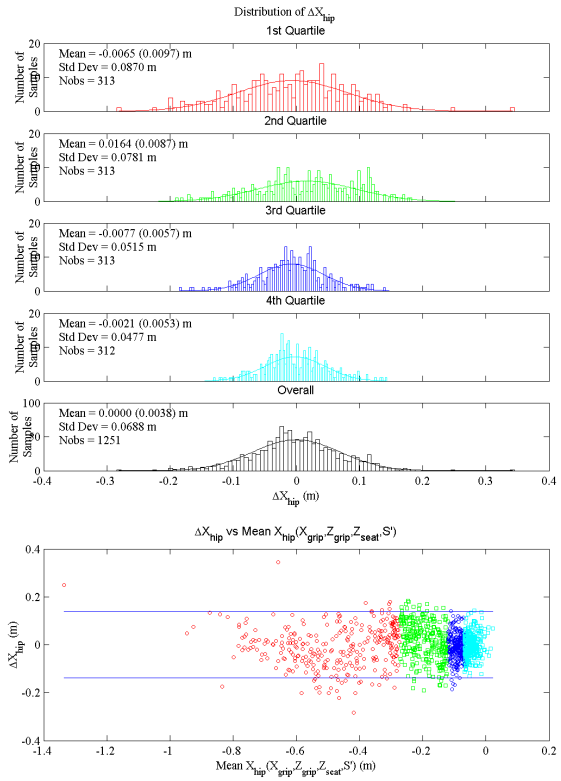

Figure 7. Distribution of the rider longitudinal hip position relative to the mean hip position.

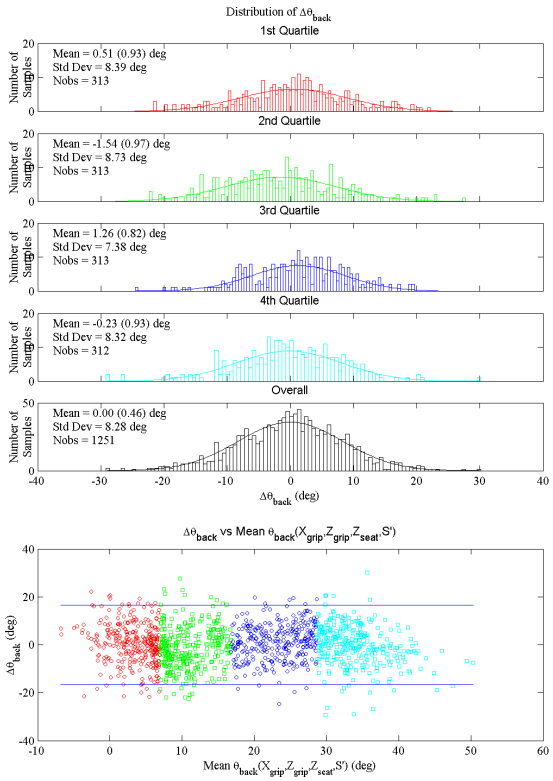

Figure 8. Distribution of the rider back angle relative to the mean back angle.

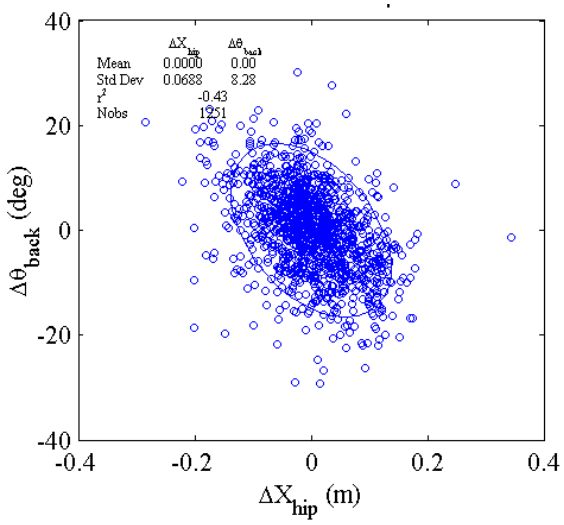

Figure 9. Variation in rider back angle vs longitudinal hip position.

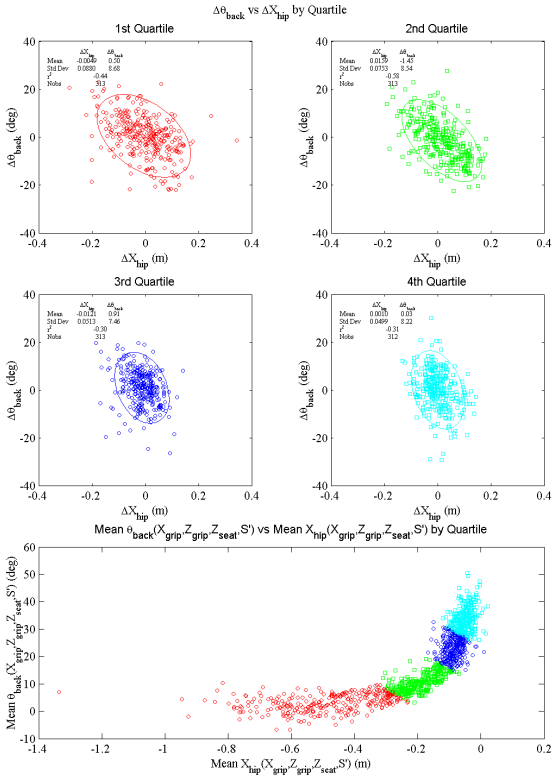

Figure 10. Variation in rider back angle vs longitudinal hip position by quartile.

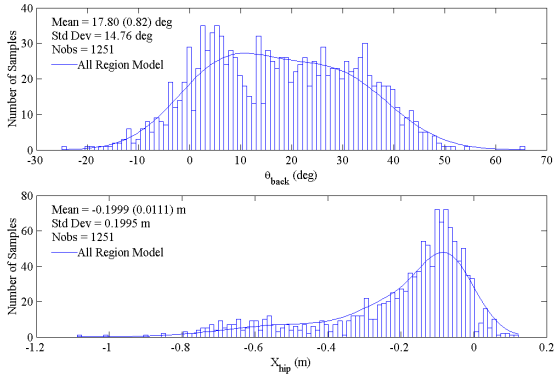

Figure 11. Modeled and observed distributions of rider longitudinal hip position and back angle.

RESULTS FOR EACH GEOGRAPHIC REGION SEPARATELY

The distribution of rider position for Japan, Europe, and the US were also assessed separately, and the resulting distribution model coefficients are also listed in Table 2. Plots illustrating the agreement between the data and the distribution models are illustrated in Appendices A, B, and C.

SOFTWARE TOOL

The descriptive models for rider position (equations (7) and (3)) and Table 2) were incorporated into a user-friendly Microsoft Excel based computer program. As illustrated in Figure 12, the MS Excel program computes rider position based on the geographic region, motorcycle seat height and hand grip position, and rider stature, and displays the results.

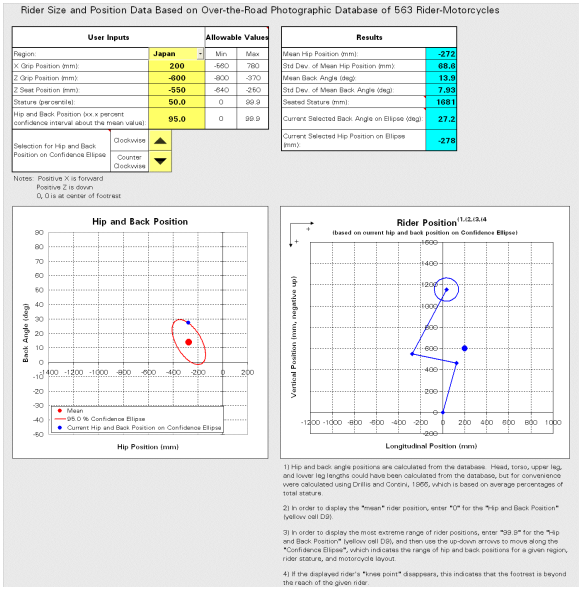

Figure 12. Rider position software tool.

CONCLUSIONS AND RECOMMENDATIONS

Descriptive statistical models for Europe, Japan, the US, and all regions were determined that describe the distribution of rider position as a function of the motorcycle-rider interface, rider stature, and geographic region. The dependent rider position variables are the longitudinal location of rider hip relative to the motorcycle footrest (x_{hip}) and rider back angle (θ_{back}). It was assumed that x_{hip} and θ_{back} are randomly distributed with normal distributions relative to mean values which are a function of the independent variables. The independent variables are the motorcycle seat height (z_{seat}) and hand grip position (x_{grip} , z_{grip}), relative to the motorcycle footrest, and the rider stature (S'). It was furthermore assumed that the standard

deviations and correlation of x_{hip} and θ_{back}, relative to the mean values, are constants. The coefficients that describe the mean, standard deviation, and correlation for each geographic region and for all geographic regions combined are listed in Table 2.

Given the motorcycle-rider interface geometry, rider stature, and geographic region, the distribution of motorcycle riding position can be estimated according to the model as follows:

1. Determine the appropriate set of coefficients listed in Table 2 to use based on the desired geographic region.
2. Calculate the mean rider position accord to the following equations:

$$\begin{aligned}\bar{x}_{hip} &= a_0 + a_1 x_{grip} + a_2 x_{grip}^2 + a_3 z_{grip} + a_4 z_{grip}^2 \\ &\quad + a_5 z_{seat} + a_6 z_{seat}^2 + a_7 S' + a_8 (S')^2 \\ \bar{\theta}_{back} &= b_0 + b_1 x_{grip} + b_2 x_{grip}^2 + b_3 z_{grip} + b_4 z_{grip}^2 \\ &\quad + b_5 z_{seat} + b_6 z_{seat}^2 + b_7 S' + b_8 (S')^2\end{aligned} \quad (7).$$

3. Calculate the approximate 95% confidence intervals for x_{hip} and θ_{back} according to the equations:

$$\begin{aligned}\bar{x}_{hip} - 1.96\, s_{x_{hip}} &\le x_{hip} \le \bar{x}_{hip} + 1.96\, s_{x_{hip}} \\ \bar{\theta}_{back} - 1.96\, s_{\theta_{back}} &\le \theta_{back} \le \bar{\theta}_{back} + 1.96\, s_{\theta_{back}}\end{aligned} \quad (8).$$

These calculations have been integrated into a Microsoft Excel computer program.

The overall agreements between the modeled and observed distributions of rider position are illustrated in Figures 11, A-11, B-11, and C-11.

It may be possible to further improve the agreement between the modeled and observed Δx_{hip} distributions by scaling or weighting the longitudinal rider position in order to model the differences in the Δx_{hip} variation that were observed in Figure 7.

It is notable that, with regard to rider back angle, the Europe and Japan models are uni-modal (with means at 24 and 13 degrees of forward lean, respectively), whereas the US model is bi-modal with peaks at 1 and 28 degrees, reflecting the different seating preferences and layouts in the data from the regions.

ACKNOWLEDGEMENTS

This work was funded by the International Motorcycle Manufacturers Association. Honda, Suzuki, and Yamaha of Japan, Harley-Davidson (US), and Yamaha (Europe) provided in-use rider position video and digitized data. DRI staff member Joe Kelly also assisted in digitizing the video data, and Brian Kebschull prepared the MS Excel computer program

REFERENCES

[1] Rogers, N.M., and Zellner, J.W., "Factors and Status of Motorcycle Airbag Feasibility Research", Paper No. 01-S9-O-207, Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Amsterdam, The Netherlands, June 4 - 7, 2001.

[2] Reed, M.P., et al., "A Statistical Method for Predicting Automobile Driving Posture", Human Factors, Vol. 44, No. 4, Winter 2002, pp 557-568.

[3] Draper, N.R., and Smith, H., Applied Regression Analysis, Second Edition, John Wiley & Sons, New York, 1981.

APPENDIX A – DISTRIBUTION OF RIDING POSITION IN JAPAN

The distribution of rider position in Japan, based on observations of 563 riders, is illustrated in Figures A-1 to A-5. Figure A-1 illustrates modeled and observed distribution of rider back angle and hip position. The modeled distributions of these dependent variables are based on the distribution of the four independent variables in Figure A.2. Figure A-3 illustrates the relationship between the rider hip position and back angle vs. the independent variables. Figure A-4 is similar to Figure A-3, but controlling for the variation in the other independent variables and illustrating the distribution model.

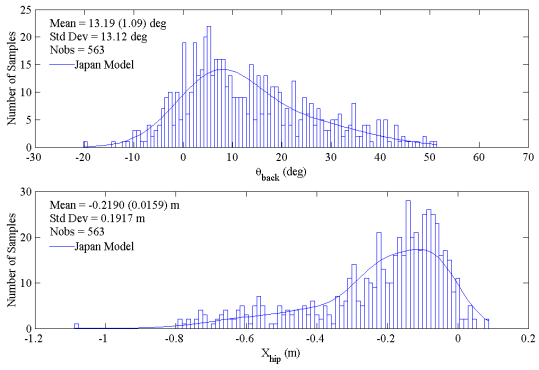

Figure A-1. Modeled and observed distribution of rider longitudinal hip position and back angle for 563 riders in Japan.

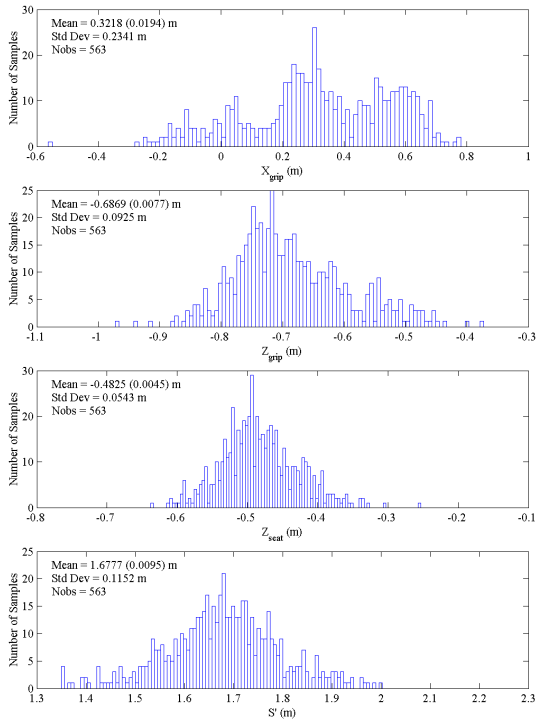

Figure A-2. Distribution of observed motorcycle seat height, hand grip location, and rider stature for 563 riders in Japan.

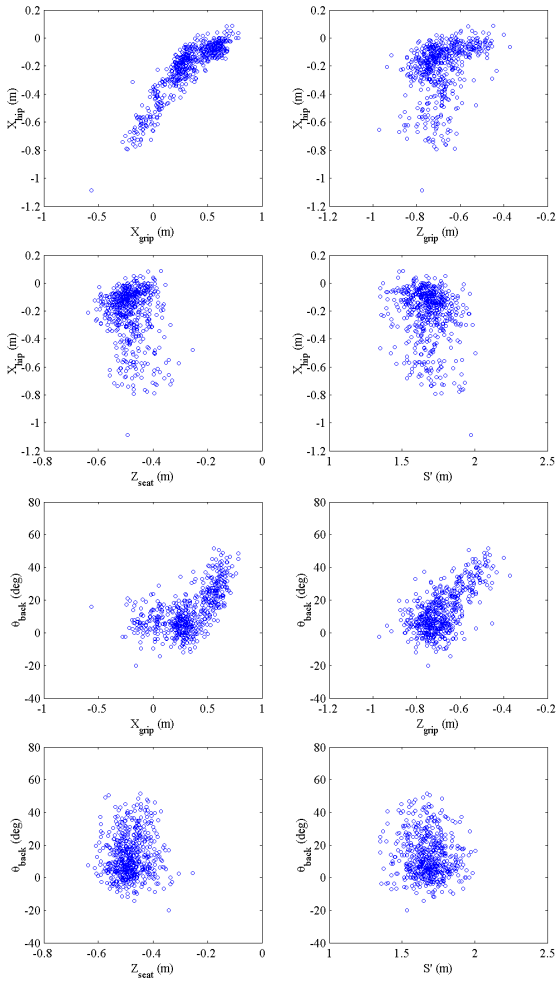

Figure A-3. Observed rider hip position and back angle vs motorcycle seat height, hand grip location, and rider stature for 563 riders in Japan.

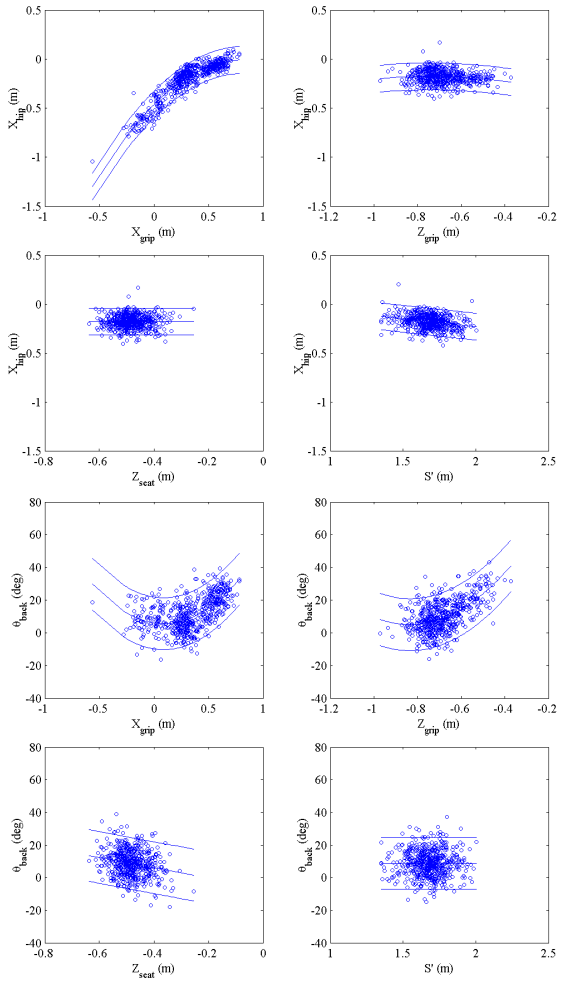

Figure A-4. Rider hip position and back angle vs. motorcycle seat height, hand grip location, and rider stature, for 563 riders in Japan, while controlling for the other independent variables.

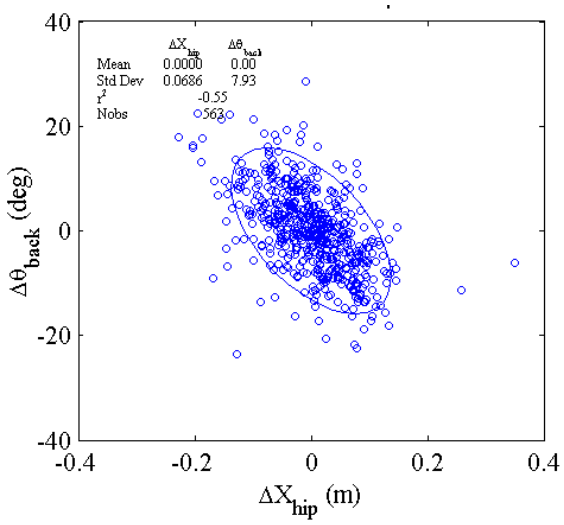

Figure A-5. Variation in rider back angle vs hip position for 563 riders in Japan.

APPENDIX B – DISTRIBUTION OF RIDING POSITION IN EUROPE

The distribution of rider position in Europe, based on observations of 501 riders in the Netherlands, is illustrated in Figures B-1 to B-5.

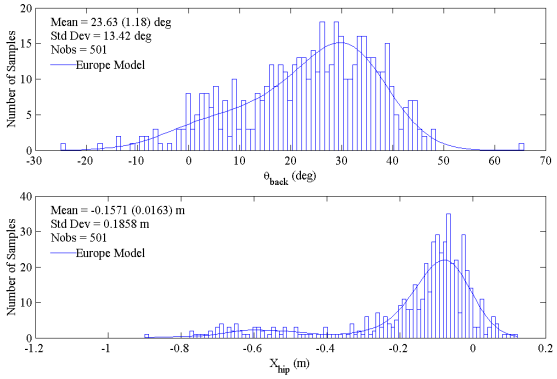

Figure B-1. Modeled and observed distribution of rider longitudinal hip position and back angle for 501 riders in Europe.

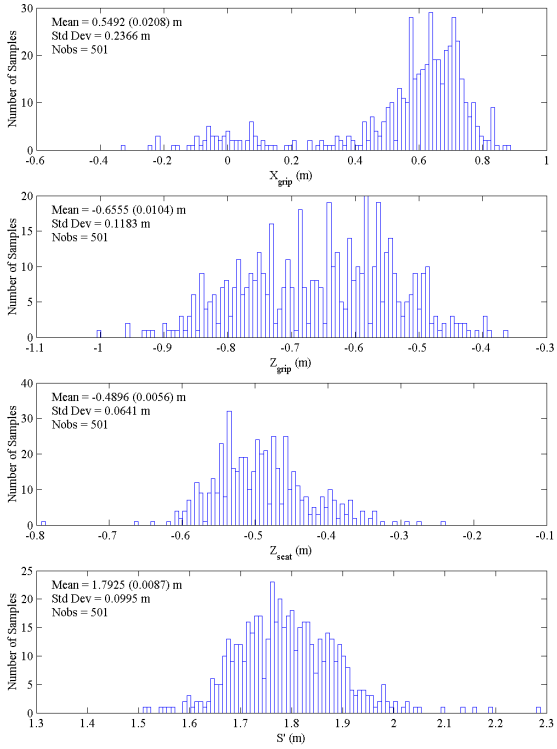

Figure B-2. Distribution of observed motorcycle seat height, hand grip location, and rider stature for 501 riders in Europe.

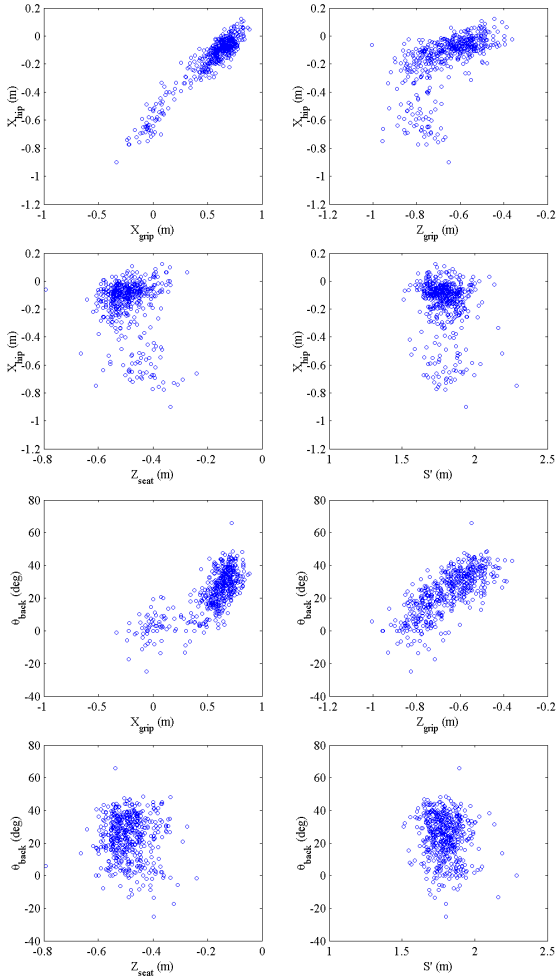

Figure B-3. Observed rider hip position and back angle vs motorcycle seat height, hand grip location, and rider stature for 501 riders in Europe.

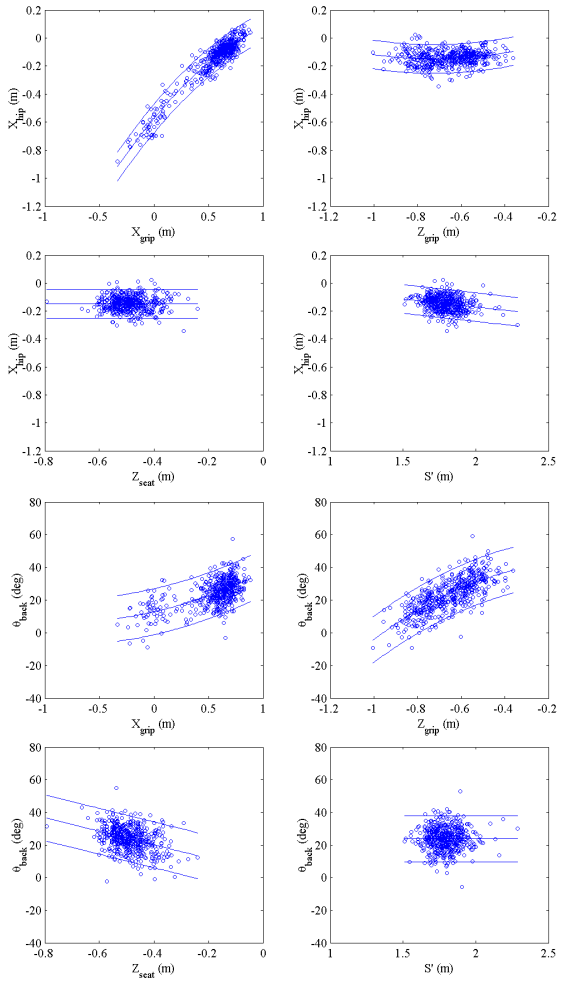

Figure B-4. Rider hip position and back angle vs. motorcycle seat height, hand grip location, and rider stature, for 501 riders in Europe, while controlling for the other independent variables.

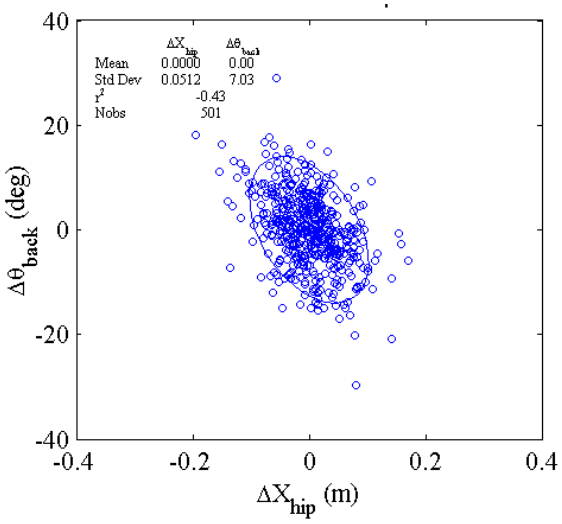

Figure B-5. Variation in rider back angle vs hip position for 501 riders in Europe.

APPENDIX C – DISTRIBUTION OF RIDING POSITION IN THE US

The distribution of rider position in the US, based on observations of 185 riders in the State of Ohio, is illustrated in Figures C-1 to C-5.

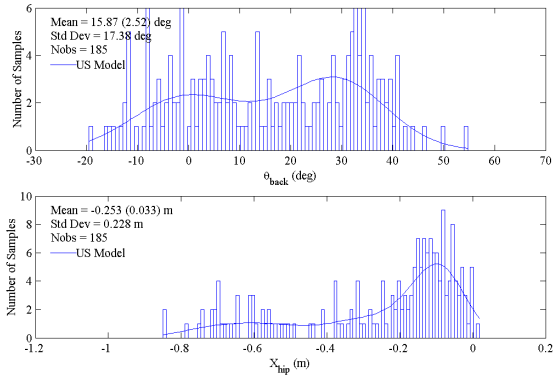

Figure C-1. Modeled and observed distribution of rider longitudinal hip position and back angle for 185 riders in the US.

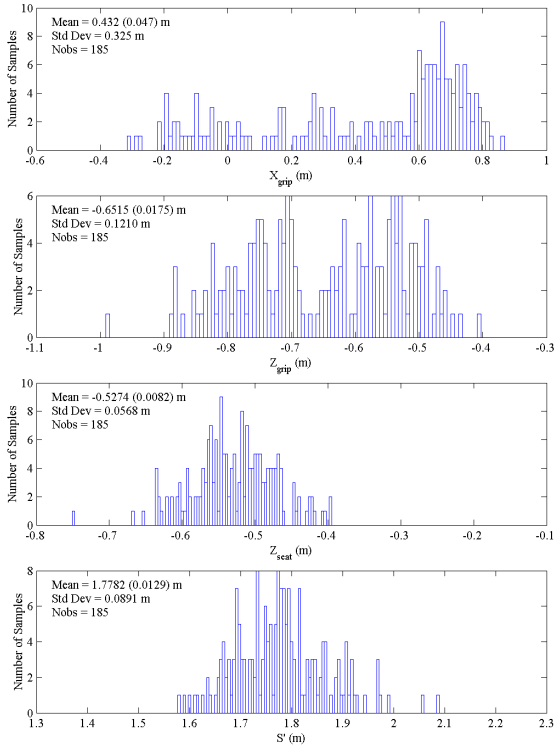

Figure C-2. Distribution of observed motorcycle seat height, hand grip location, and rider stature for 185 riders in the US.

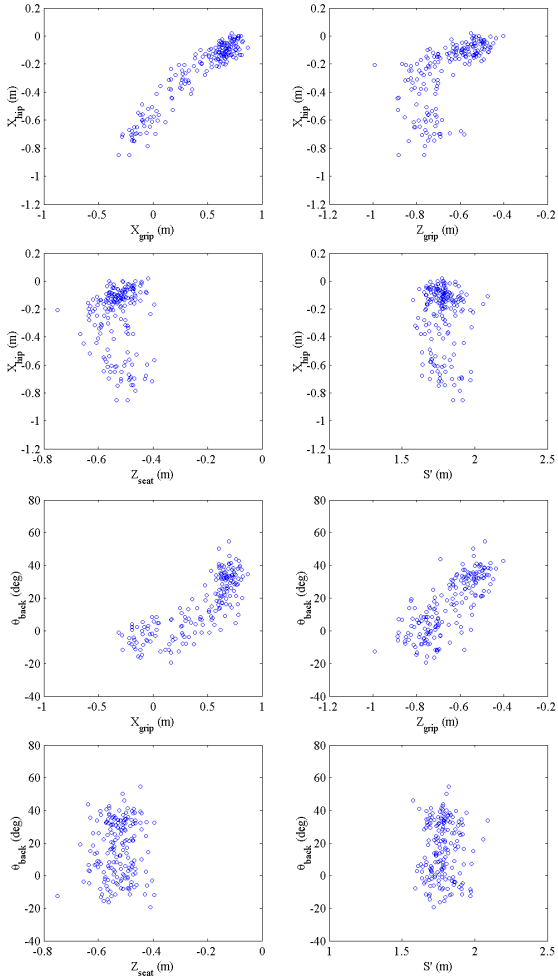

Figure C-3. Observed rider hip position and back angle vs motorcycle seat height, hand grip location, and rider stature for 185 riders in the US.

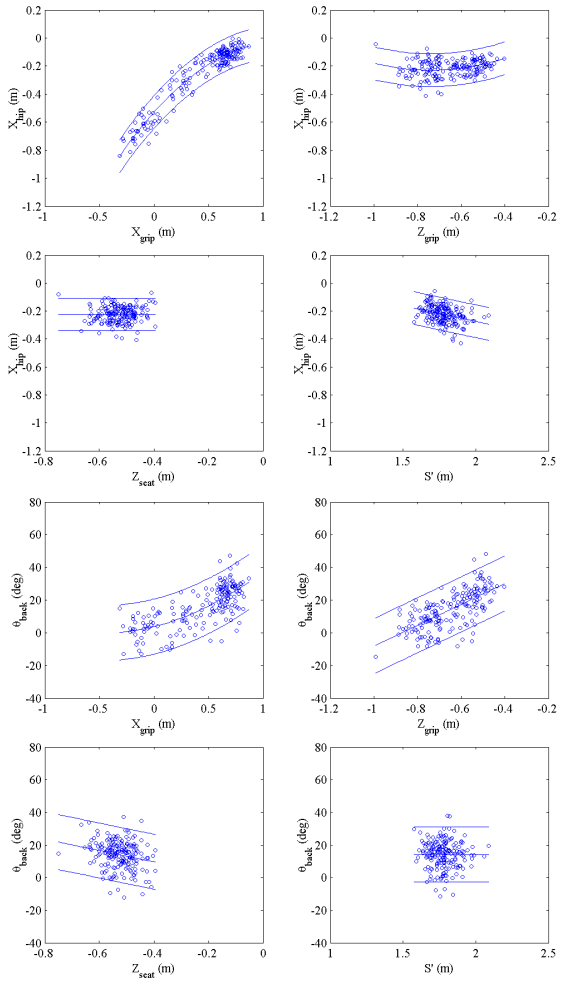

Figure C-4. Rider hip position and back angle vs. motorcycle seat height, hand grip location, and rider stature, for 185 riders in the US, while controlling for the other independent variables.

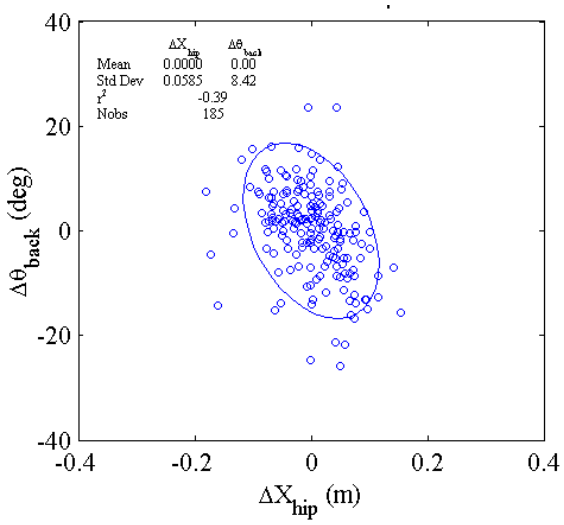

Figure C-5. Variation in rider back angle vs hip position for 185 riders in the US.

APPENDIX D

The RPRED statistic was the figure of merit used to select the regression models. It is based on the PRESS statistic described in Section 6.8 of Draper and Smith [3]. The RPRED statistic can be calculated according to the following equation:

$$RPRED = 1 - \frac{\text{PRESS}}{\text{SS}_{TOT}}$$

Where

$$PRESS = \sum_{k=1}^{n} \left(y_k - \hat{y}_{k,-k} \right)^2$$

$$SS_{TOT} = \sum_{k=1}^{n} \left(y_k - \bar{y} \right)^2$$

$$\bar{y} = \frac{1}{n} \sum_{k=1}^{n} y_k$$

The RPRED statistic is similar to the regression model R^2 statistic, except that PRESS residuals are used instead of ordinary residuals. Whereas ordinary residuals are the difference between the observed value for y and the estimated value $\hat{y}$, PRESS residuals are the difference between the observed y and $\hat{y}$ predicted by a model in which one rating at a time had been set aside and not used to identify the model. Therefore RPRED is a measure of both the fit and the predictive capabilities, and RPRED values approaching 1 are desirable.