Let's Interact!
Modeling Interaction Effects in Linear and Generalized Linear Models using SAS®

Timothy B. Gravelle
Principal Scientist & Director, Insights Lab
PriceMetrix Inc.

Certified Base Programmer for SAS 9

Certified Statistical Business Analyst Using SAS®9: Regression and Modeling

Which of these statements are accurate?

The REG Procedure

Dependent Variable: TOTAL_REV_11

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	739.79385	7.74079	95.57	<.0001
ACCTS_PER_HH	46.24691	22.25271	2.08	0.0379
MEDIAN_HH_ASSETS	-0.30548	0.12804	-2.39	0.0172
ACCTS_PER_HH_MED_HH_ASSETS	0.56231	0.17897	3.14	0.0017

1. The coefficient for ACCTS_PER_HH is statistically significant. Therefore, the hypothesis of a relationship between ACCTS_PER_HH and TOTAL_REV_11 (the dependent variable) is confirmed.
2. The coefficient for MEDIAN_HH_ASSETS is statistically significant. Therefore, the hypothesis of a relationship between MEDIAN_HH_ASSETS and TOTAL_REV_11 is confirmed.
3. The coefficient for the interaction term ACCTS_PER_HH_MEDIAN_HH_ASSETS is statistically significant. Therefore, the hypothesis relating the combination of ACCTS_PER_HH and MEDIAN_HH_ASSETS to TOTAL_REV_11 is confirmed.

Topics

- A (brief) review of the theoretical framework for interaction effects
- Data preparation for testing interactions
- Specifying interactive models in PROC REG and PROC LOGISTIC
- Graphical displays for interaction effects

Two examples

- Predicting investment advisors' future productivity (linear model)
- Canadian attitudes toward Canada-US relations (ordinal logit model)

Theory

He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may be cast.

– Leonardo da Vinci

PriceMetrix®

REALIZE YOUR VALUE

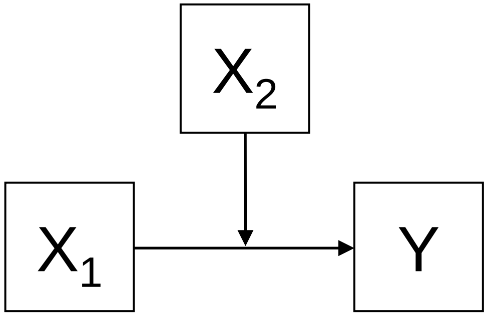
Theorizing and specifying interaction effects

- A “system” comprising 3 variables (Jaccard and Turrisi 2003; Jaccard and Dodge 2004):
  - Dependent variable
  - “Focal” independent variable
  - Moderator variable
- Example: employment income (DV), education (focal IV) and sex (moderator)

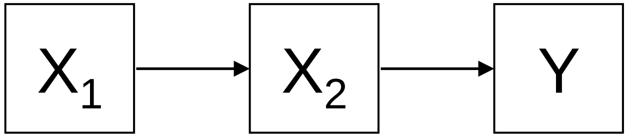
Theorizing and specifying interaction effects

- Moderation vs. mediation (Baron and Kenny 1986):
  - A moderator variable "affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable."
  - A mediator variable "accounts for the relation between the predictor and the criterion."

Moderation



Mediation



PriceMetrix
REALIZE YOUR VALUE

Theorizing and specifying interaction effects

- $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$

$\hat{Y}$	predicted value (conditional mean) of the dependent variable
β_0	intercept term
β_1 β_2	coefficients for the independent variables; "lower-order" terms
β_{12}	coefficient for the interaction term; "higher order" term
X_1 X_2	values taken by the independent variables

PriceMetrix®
REALIZE YOUR VALUE

Theorizing and specifying interaction effects

- When testing an interaction effect, the lower-order terms (β_1 and β_2) must still be present in the model. Otherwise, the model is not "hierarchically well-formulated."
- Even when included in the model β_1 and β_2 are not of primary interest. And they are **not** interpreted as the "main effects" of β_1 and β_2 or the "effects of β_1 and β_2 in general."
- Rather, they indicate the effect of X_1 on $\hat{Y}$ when X_2 is 0, and the effect of X_2 on $\hat{Y}$ when X_1 is 0.

PriceMetrix

REALIZE YOUR VALUE

Case #1

Predicting Investment Advisor Productivity

PriceMetrix®

REALIZE YOUR VALUE

Predicting Investment Advisor Productivity

- How can one predict investment advisors' future productivity (revenue) given data on advisors' performance in a baseline period, personal characteristics and information on the advisors' books?
- Data are drawn from the proprietary PriceMetrix retail wealth management database.
  - 1,010 investment advisors
  - Multiple firms across North America (Canada and US)
  - Advisors with 5 to 20 years of industry experience as of end-of-year 2006

Predicting Investment Advisor Productivity

Advisor Data: Basic Descriptive Statistics

The MEANS Procedure

Variable	Mean	Minimum	Maximum	Median	Std Dev	N
TOTAL_REV_11	760.17	14.61	3888.97	630.72	481.83	1010
TRANS_REV_06	327.19	7.40	2184.26	253.86	273.46	1010
TRAILER_REV_06	101.43	0.00	726.07	77.91	86.99	1010
FEE_REV_06	199.22	0.00	2443.41	101.16	257.27	1010
EXPERIENCE_YEARS	11.96	5.05	20.00	11.81	4.27	1010
TEAM	0.06	0.00	1.00	0.00	0.24	1010
CORE_HH_COUNT	74.17	1.00	383.00	65.00	44.75	1010
SMALL_HH_COUNT	173.33	4.00	1269.00	142.00	134.40	1010
RETIREMENT_ACCT_COUNT	127.65	0.00	1141.00	112.00	102.08	1010
ACCTS_PER_HH	2.08	1.07	4.01	2.08	0.40	1010
MEDIAN_HH_ASSETS	150.71	0.10	715.06	139.55	87.65	1010

PriceMetrix
REALIZE YOUR VALUE

Data Preparation

Data Preparation

- All of the standard assumptions underpinning regression analysis continue to apply.
  - linearity in the predictors
  - normality
  - constant error variance (homoscedasticity)
  - independence of the errors
  - absence of high collinearity among the predictors

Mean Centering

- One technique that is especially relevant when modeling interaction effects is *mean-centering*.
- Involves subtracting the mean from the original scores, resulting in new scores with a mean of zero.
- Zero now has the interpretation of a variable's pre-transformation mean value.
- Jaccard and Turrisi recommend this strategy as a way to "force the coefficients to reflect parameters that are of theoretical interest" (2003: 15).
- *Median*-centering is also an option.

PriceMetrix
REALIZE YOUR VALUE

Mean Centering

- Easy to implement in SAS using PROC STDIZE using the METHOD=MEAN option (default method creates standardized (Z) scores).
- Example:

```
PROC STDIZE DATA=data_1 OUT=data_2 METHOD=MEAN;
VAR TRANS_REV_06 TRAILER_REV_06 FEE_REV_06
  EXPERIENCE_YEARS CORE_HH_COUNT
  SMALL_HH_COUNT RETIREMENT_ACCT_COUNT
  ACCTS_PER_HH MEDIAN_HH_ASSETS;
RUN;
```

Product Terms

- Once independent variables are mean centered, product terms for interaction effects can be created in a data step.
- Unlike PROC GLM, interaction terms cannot be entered directly into PROC REG.
- Example:

```
DATA data_2;
SET data_2;
ACCTS_PER_HH_MED_HH_ASSETS=
  ACCTS_PER_HH*MEDIAN_HH_ASSETS;
RUN;
```

Let's Interact!
Specifying the Model with PROC REG

PriceMetrix®
REALIZE YOUR VALUE

Model Specification

```
PROC REG DATA=data_2 OUTEST=parmest;
MAINEFFECTS: MODEL TOTAL_REV_11 = TRANS_REV_06
  TRAILER_REV_06 FEE_REV_06 EXPERIENCE_YEARS TEAM
  CORE_HH_COUNT SMALL_HH_COUNT RETIREMENT_ACCT_COUNT
  ACCTS_PER_HH MEDIAN_HH_ASSETS
  /ADJRSQ CLB STB VIF;
INTERACTION: MODEL TOTAL_REV_11 = TRANS_REV_06
  TRAILER_REV_06 FEE_REV_06 EXPERIENCE_YEARS TEAM
  CORE_HH_COUNT SMALL_HH_COUNT RETIREMENT_ACCT_COUNT
  ACCTS_PER_HH MEDIAN_HH_ASSETS
  ACCTS_PER_HH_MED_HH_ASSETS
  /ADJRSQ CLB STB VIF;
INT_EFFECT: TEST ACCTS_PER_HH_MED_HH_ASSETS=0;
RUN; QUIT;
```

PriceMetrix
REALIZE YOUR VALUE

Results: Main Effects Model

The REG Procedure
Model: MAINEFFECTS
Dependent Variable: TOTAL_REV_11

Number of Observations Read	1010
Number of Observations Used	1010

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	190238759	19023876	431.86	<.0001
Error	999	44007296	44051		
Corrected Total	1009	234246055			

Root MSE	209.88413	R-Square	0.8121
Dependent Mean	760.17482	Adj R-Sq	0.8103
Coeff Var	27.60998		

PriceMetrix®
REALIZE YOUR VALUE

Results: Main Effects Model

Parameter Estimates

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	751.26991	6.85512	109.59	<.0001
TRANS_REV_06	0.82323	0.02916	28.24	<.0001
TRAILER_REV_06	0.77207	0.09913	7.79	<.0001
FEE_REV_06	0.97834	0.03272	29.90	<.0001
EXPERIENCE_YEARS	-12.56149	1.62893	-7.71	<.0001
TEAM	142.76120	29.46273	4.85	<.0001
CORE_HH_COUNT	1.32491	0.27042	4.90	<.0001
SMALL_HH_COUNT	-0.15372	0.10117	-1.52	0.1290
RETIREMENT_ACCT_COUNT	0.47258	0.11506	4.11	<.0001
ACCTS_PER_HH	53.81291	22.22005	2.42	0.0156
MEDIAN_HH_ASSETS	-0.11352	0.11302	-1.00	0.3154

PriceMetrix
REALIZE YOUR VALUE

Results: Interactive Model

The REG Procedure
Model: INTERACTION
Dependent Variable: TOTAL_REV_11

Number of Observations Read	1010
Number of Observations Used	1010

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	190669800	17333618	396.98	<.0001
Error	998	43576255	43664		
Corrected Total	1009	234246055			

Root MSE	208.95833	R-Square	0.8140
Dependent Mean	760.17482	Adj R-Sq	0.8119
Coeff Var	27.48819		

PriceMetrix®
REALIZE YOUR VALUE

Results: Interactive Model

Parameter Estimates

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	739.79385	7.74079	95.57	<.0001
TRANS_REV_06	0.80831	0.02941	27.48	<.0001
TRAILER_REV_06	0.76138	0.09875	7.71	<.0001
FEE_REV_06	0.96004	0.03309	29.01	<.0001
EXPERIENCE_YEARS	-12.66608	1.62208	-7.81	<.0001
TEAM	140.36661	29.34267	4.78	<.0001
CORE_HH_COUNT	1.64718	0.28810	5.72	<.0001
SMALL_HH_COUNT	-0.26999	0.10731	-2.52	0.0120
RETIREMENT_ACCT_COUNT	0.50773	0.11510	4.41	<.0001
ACCTS_PER_HH	46.24691	22.25271	2.08	0.0379
MEDIAN_HH_ASSETS	-0.30548	0.12804	-2.39	0.0172
ACCTS_PER_HH_MED_HH_ASSETS	0.56231	0.17897	3.14	0.0017

PriceMetrix
REALIZE YOUR VALUE

Results: Interactive Model

Test INT_EFFECT Results for Dependent

Variable TOTAL_REV_11

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	431040	9.87	0.0017
Denominator	998	43664		

PriceMetrix®
REALIZE YOUR VALUE

Results: Summary Dataset

Partial output from the parmest dataset:

MODEL	_TYPE_	_DEPVAR_	Intercept	ACCTS_ PER_HH	MEDIAN_ HH_ASSETS	ACCTS_PER_ HH_MED_HH_ ASSETS	_RSQ_
MAINEFFECTS	PARMS	TOTAL_REV_11	751.270	53.813	-0.114	.	0.812
INTERACTION	PARMS	TOTAL_REV_11	739.794	46.247	-0.305	0.562	0.814

PriceMetrix®
REALIZE YOUR VALUE

A Plot is Worth a Thousand Words (or Coefficients)

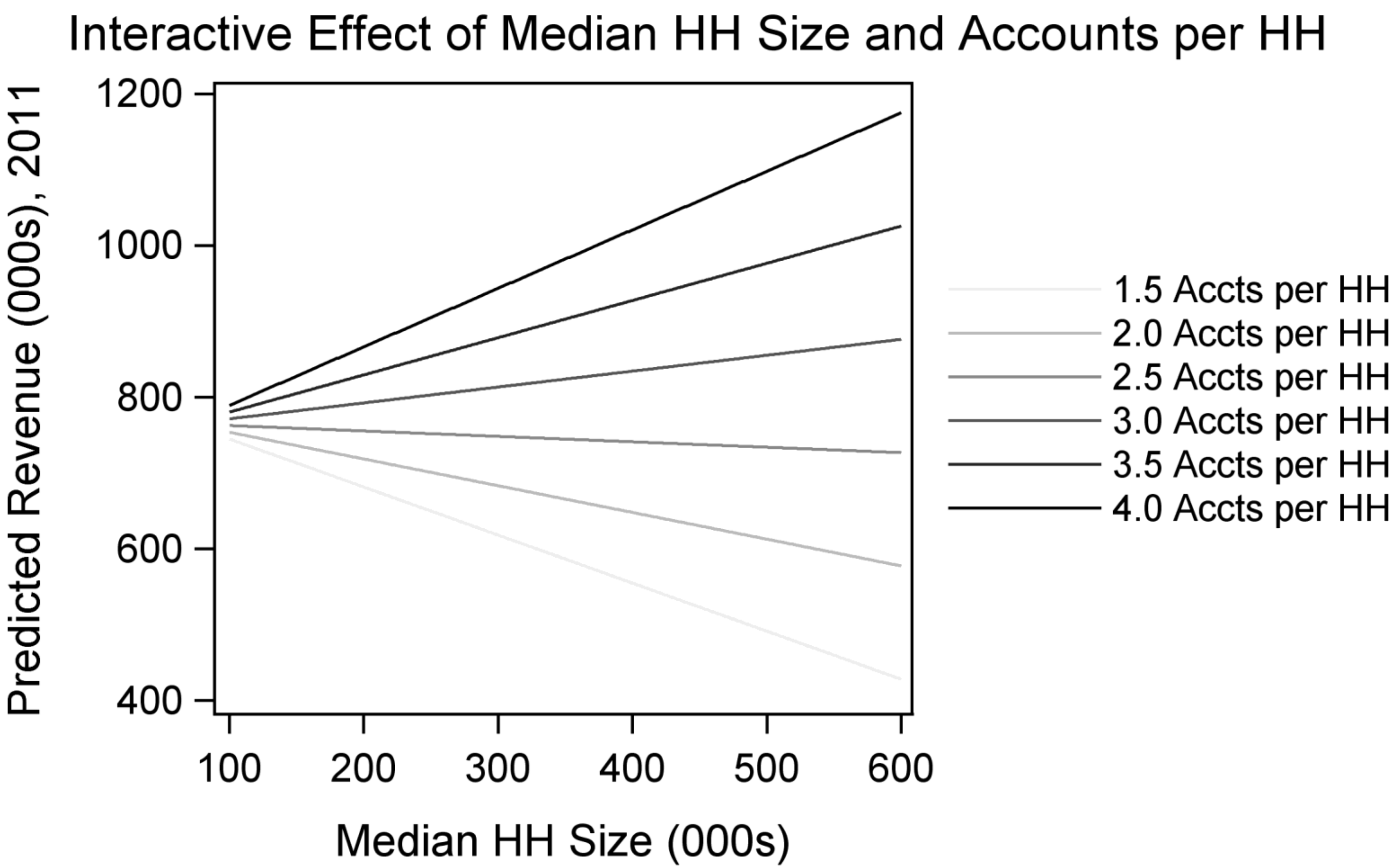
PriceMetrix®
REALIZE YOUR VALUE

Graphical Depictions of Interaction Effects

- Two strategies:
  - **Effect plots** (effect displays) depict the strength and direction of the relationship between the focal independent variable and dependent variable at different levels of the moderator variable.
  - **Coefficient plots** display the coefficient (and confidence interval) for the focal independent variable with the scores for the moderator variable centered at different values. This serves to highlight the regions of significance of the focal independent variable.

PriceMetrix®
REALIZE YOUR VALUE

Effect Plot


Interactive Effect of Median HH Size and Accounts per HH
Predicted Revenue (000s), 2011
1200
1000
800
600
400
100
200
300
400
500
600
Median HH Size (000s)
1.5 Accts per HH
2.0 Accts per HH
2.5 Accts per HH
3.0 Accts per HH
3.5 Accts per HH
4.0 Accts per HH

Effect Plot

1. Output the means of the variables involved in the interaction and create macro variables (PROC UNIVARIATE and CALL SYMPUT).
2. Use nested DO loops in a DATA STEP to generate the desired values of the variables involved in the interaction and multiply these values out by the model coefficients (using the dataset created by the PROC REG outest option).
3. In a DATA step, restructure the resulting dataset (one row for each value of the focal IV; multiple columns for different values of the moderator).
4. Plot the predicted values of the DV (PROC SGPLOT).

Effect Plot

```
ODS OUTPUT BasicMeasures=means;
PROC UNIVARIATE DATA=data_1;
VAR ACCTS_PER_HH MEDIAN_HH_ASSETS;
RUN;

DATA _NULL_;
SET means;
IF VarName="ACCTS_PER_HH" AND LocMeasure="Mean"
   THEN CALL SYMPUT('AVG_ACCTS_PER_HH', LocValue);
IF VarName="MEDIAN_HH_ASSETS" AND LocMeasure="Mean"
   THEN CALL SYMPUT('AVG_MEDIAN_HH_ASSETS', LocValue);
RUN;
```

Effect Plot

```
DATA plot_1 (DROP=i j _MODEL_);
SET parmest (WHERE=(_MODEL_="INTERACTION") KEEP=_MODEL_ Intercept
   MEDIAN_HH_ASSETS ACCTS_PER_HH ACCTS_PER_HH_MED_HH_ASSETS
   RENAME=(MEDIAN_HH_ASSETS=b_MED_ASSETS ACCTS_PER_HH=b_ACCTS
   ACCTS_PER_HH_MED_HH_ASSETS=b_MED_ASSETS_ACCTS));
DO i=100 TO 600;
   DO j=1.5 TO 4 BY 0.5;
      MEDIAN_HH_ASSETS=i;
      MEDIAN_HH_ASSETS_CTR=i - INPUT(&AVG_MEDIAN_HH_ASSETS, BEST12.);
      ACCTS_PER_HH=j;
      ACCTS_PER_HH_CTR=j - INPUT(&AVG_ACCTS_PER_HH, BEST12.);
      PRED=Intercept +                                    /* Intercept */
      (b_MED_ASSETS * MEDIAN_HH_ASSETS_CTR) +    /* Median HH Assets */
      (b_ACCTS * ACCTS_PER_HH_CTR) +       /* Accounts per household */
      (b_MED_ASSETS_ACCTS * (MEDIAN_HH_ASSETS_CTR * ACCTS_PER_HH_CTR))
                                                        /* Interaction */
      ;
      OUTPUT;
      END;
   END;
RUN;
```

PriceMetrix
REALIZE YOUR VALUE

Effect Plot

```
DATA plot_2;
MERGE  plot_1 (WHERE=(ACCTS_PER_HH=1.5) RENAME=(PRED=PRED_1_5))
       plot_1 (WHERE=(ACCTS_PER_HH=2.0) RENAME=(PRED=PRED_2_0))
       plot_1 (WHERE=(ACCTS_PER_HH=2.5) RENAME=(PRED=PRED_2_5))
       plot_1 (WHERE=(ACCTS_PER_HH=3.0) RENAME=(PRED=PRED_3_0))
       plot_1 (WHERE=(ACCTS_PER_HH=3.5) RENAME=(PRED=PRED_3_5))
       plot_1 (WHERE=(ACCTS_PER_HH=4.0) RENAME=(PRED=PRED_4_0));
BY MEDIAN_HH_ASSETS;
RUN;
```

Effect Plot

```
ODS GRAPHICS ON /BORDER=OFF HEIGHT=2.5IN WIDTH=4IN;
ODS LISTING IMAGE_DPI=600 STYLE=JOURNAL SGE=OFF;

PROC SGPLOT DATA=plot_2;
TITLE "Interactive Effect of Median HH Size and Accounts per HH";
SERIES Y=PRED_1_5 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CXE1E1E1)
   LEGENDLABEL="1.5 Accts per HH";
SERIES Y=PRED_2_0 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CXB4B4B4)
   LEGENDLABEL="2.0 Accts per HH";
SERIES Y=PRED_2_5 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CX878787)
   LEGENDLABEL="2.5 Accts per HH";
SERIES Y=PRED_3_0 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CX5A5A5A)
   LEGENDLABEL="3.0 Accts per HH";
```

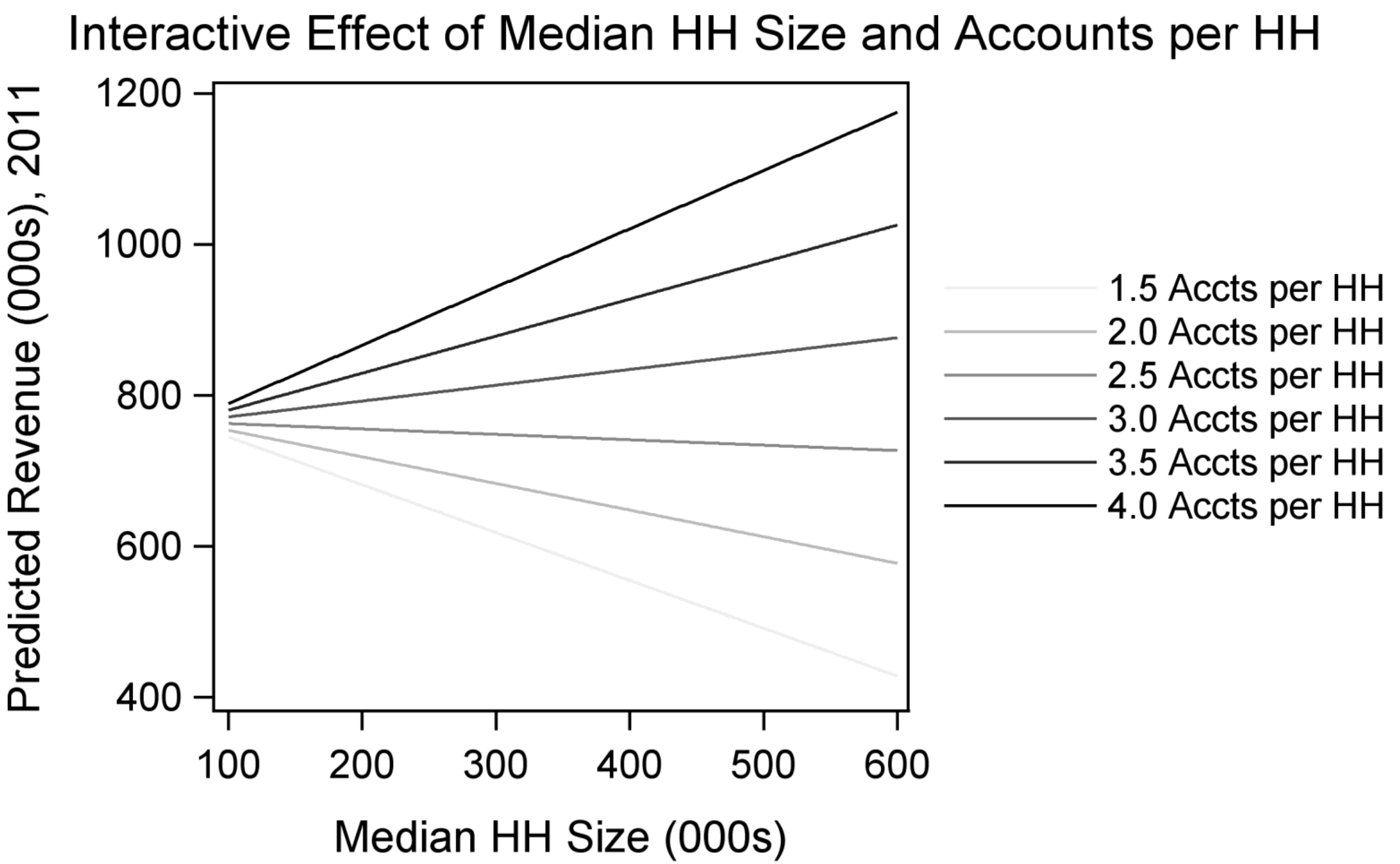
Effect Plot

```
SERIES Y=PRED_3_5 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CX2D2D2D)
   LEGENDLABEL="3.5 Accts per HH";
SERIES Y=PRED_4_0 X=MEDIAN_HH_ASSETS
   /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=CX000000)
   LEGENDLABEL="4.0 Accts per HH";
KEYLEGEND /POSITION=RIGHT
   LOCATION=OUTSIDE ACROSS=1 DOWN=6 NOBORDER;
YAXIS MIN=400 MAX=1200 VALUES=(400 600 800 1000 1200)
   OFFSETMIN=0.02
   LABEL="Predicted Revenue (000s), 2011";
XAXIS  MIN=100 MAX=600 VALUES=(100 200 300 400 500 600)
   OFFSETMIN=0.02
   LABEL="Median HH Size (000s)";
RUN;

ODS GRAPHICS OFF;
```

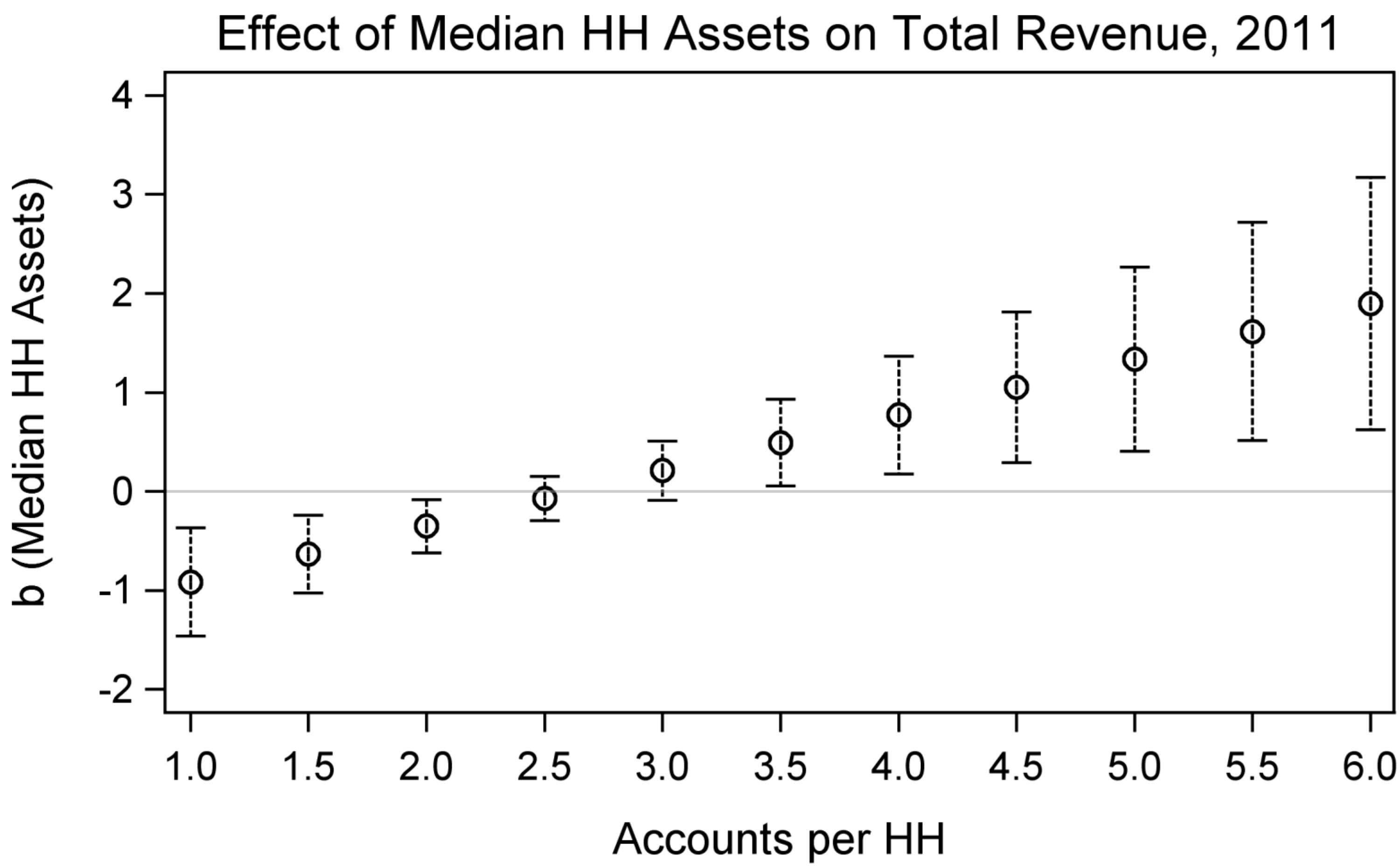
PriceMetrix
REALIZE YOUR VALUE

Effect Plot



PriceMetrix
REALIZE YOUR VALUE

Coefficient Plot



PriceMetrix®
REALIZE YOUR VALUE

Coefficient Plot

1. Mean-center all of the continuous independent variables, except for the variables involved in the interaction (PROC STDIZE)
2. In a DATA step, create an empty dataset (zero observations) that will contain the model coefficients, confidence intervals and the centering value of the moderator.

PriceMetrix®
REALIZE YOUR VALUE

Coefficient Plot

3. Run the %INTPROBE macro, iteratively performing the following steps using a macro %DO loop:
    1. In a DATA step, increment the value of the moderator, calculate the new (centred) moderator and the interaction product term.
    2. Run the regression model (PROC REG) and output the coefficients using the ODS OUTPUT statement.
    3. In another DATA step, add the model coefficients to the coefficients dataset.
    4. Delete the datasets created within the iteration of the macro.

Coefficient Plot

```
PROC STDIZE DATA=data_1 OUT=data_3 METHOD=MEAN;
VAR TRANS_REV_06 TRAILER_REV_06 FEE_REV_06
   EXPERIENCE_YEARS CORE_HH_COUNT SMALL_HH_COUNT
   RETIREMENT_ACCT_COUNT;
RUN;
DATA parmsint;

LENGTH Variable $50 ACCTS_PER_HH_CENTER 3.
   Estimate LCL UCL 8.;
FORMAT p PVALUE6.4;
RUN;
```

PriceMetrix®
REALIZE YOUR VALUE

Coefficient Plot

```
%MACRO INTPROBE (DataIn=, DataOut= );
   %DO ACCT_HH=10 %TO 60;
      DATA &DataOut.;
      SET &DataIn.;
      CENTRE_VALUE=ROUND((&ACCT_HH*0.1), 0.1);
      ACCTS_PER_HH=ACCTS_PER_HH - CENTRE_VALUE ;
      ACCTS_PER_HH_MED_HH_ASSETS=ACCTS_PER_HH*MEDIAN_HH_ASSETS;
      RUN;

      PROC REG DATA=&DataOut.;
      MODEL TOTAL_REV_11 = TRANS_REV_06 TRAILER_REV_06 FEE_REV_06
      EXPERIENCE_YEARS TEAM CORE_HH_COUNT SMALL_HH_COUNT
      RETIREMENT_ACCT_COUNT ACCTS_PER_HH MEDIAN_HH_ASSETS
      ACCTS_PER_HH_MED_HH_ASSETS
      /CLB STB;
      ODS OUTPUT ParameterEstimates=parms;
      RUN; QUIT;
```

PriceMetrix
REALIZE YOUR VALUE

Coefficient Plot

```
    DATA parmsctr (KEEP=Variable Estimate StandardizedEst
       LowerCL UpperCL Probt ACCTS_PER_HH_CENTER
       RENAME=(StandardizedEst=StdCoeff LowerCL=LCL UpperCL=UCL
       Probt=p));
    LENGTH Variable $50 ACCTS_PER_HH_CENTER 3.;
    SET parms;
    ACCTS_PER_HH_CENTER=ROUND((&ACCT_HH*0.1), 0.1);
    FORMAT ACCTS_PER_HH_CENTER 3.1;
    RUN;
    DATA parmsint;
    SET parmsint parmsctr;
    IF Variable="" THEN DELETE;
    FORMAT ACCTS_PER_HH_CENTER 3.1;
    RUN;
    PROC DATASETS LIB=work NOLIST;
    DELETE &DataOut. parms parmsctr;
    RUN; QUIT;
 %END;
%MEND INTPROBE;
```

PriceMetrix
REALIZE YOUR VALUE

Coefficient Plot

```
%INTPROBE(DataIn=data_3, DataOut=data_4);

PROC SORT DATA=parmsint;
BY Variable;
RUN;
```

PriceMetrix®
REALIZE YOUR VALUE

Coefficient Plot

Variable	ACCTS_PER_ HH_CENTER	Parameter Estimate	Lower 95% CL Parameter	Upper 95% CL Parameter	Pr > |t|
MEDIAN_HH_ASSETS	1.0	-0.915	-1.462	-0.368	0.0011
MEDIAN_HH_ASSETS	1.1	-0.859	-1.374	-0.344	0.0011
MEDIAN_HH_ASSETS	1.2	-0.803	-1.286	-0.319	0.0012
. . .					
MEDIAN_HH_ASSETS	1.9	-0.409	-0.697	-0.121	0.0054
MEDIAN_HH_ASSETS	2.0	-0.353	-0.619	-0.086	0.0096
MEDIAN_HH_ASSETS	2.1	-0.296	-0.545	-0.048	0.0195
MEDIAN_HH_ASSETS	2.2	-0.240	-0.475	-0.006	0.0447
MEDIAN_HH_ASSETS	2.3	-0.184	-0.409	0.041	0.1091
MEDIAN_HH_ASSETS	2.4	-0.128	-0.349	0.093	0.2569
MEDIAN_HH_ASSETS	2.5	-0.072	-0.294	0.151	0.5281
. . .					
MEDIAN_HH_ASSETS	3.0	0.210	-0.090	0.509	0.1694
MEDIAN_HH_ASSETS	3.1	0.266	-0.058	0.590	0.1075
MEDIAN_HH_ASSETS	3.2	0.322	-0.028	0.673	0.0716
MEDIAN_HH_ASSETS	3.3	0.378	0.000	0.757	0.0500
MEDIAN_HH_ASSETS	3.4	0.435	0.027	0.842	0.0366
MEDIAN_HH_ASSETS	3.5	0.491	0.054	0.928	0.0279
MEDIAN_HH_ASSETS	3.6	0.547	0.079	1.015	0.0220
. . .					

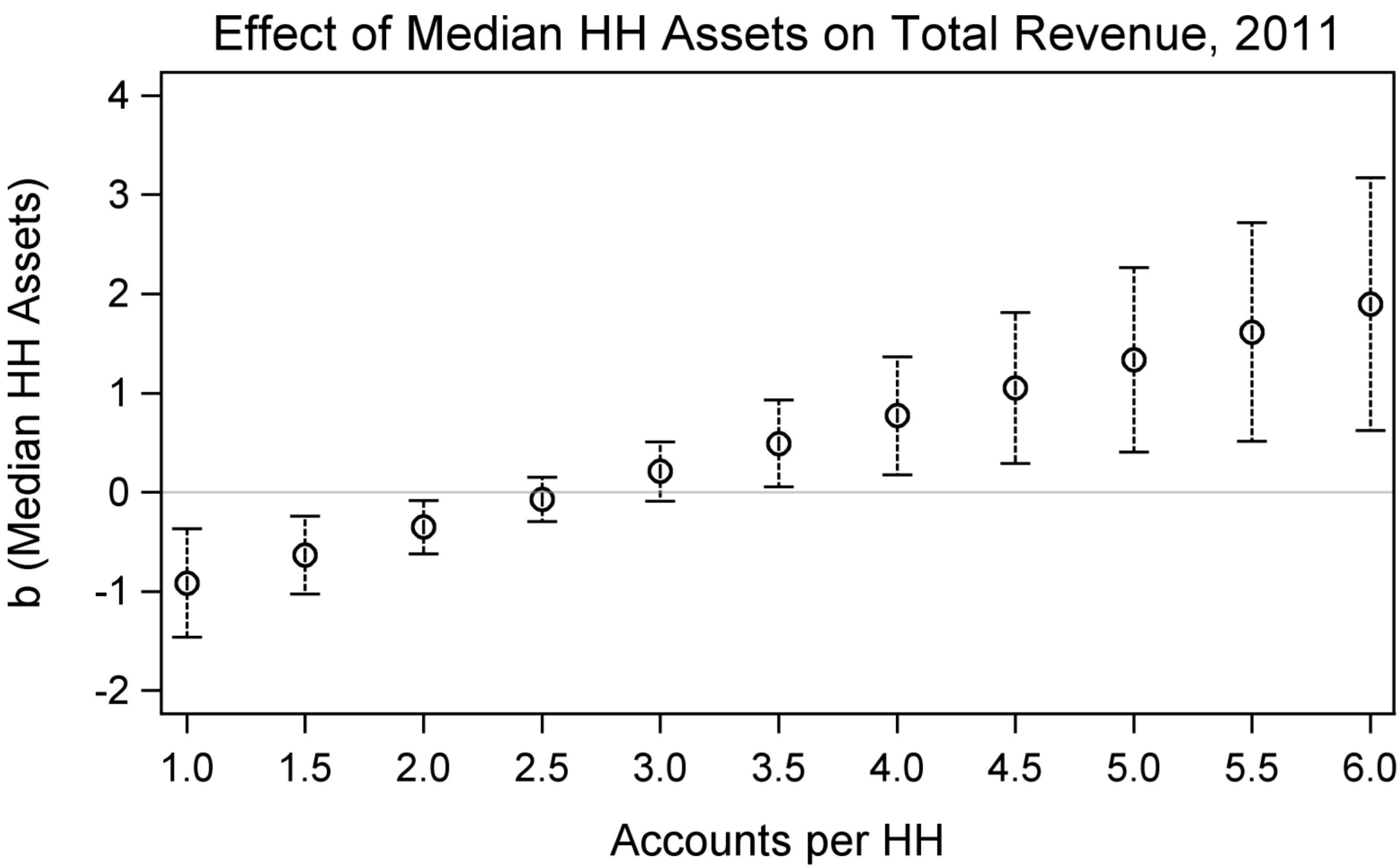
Coefficient Plot

```
ODS GRAPHICS ON /BORDER=OFF HEIGHT=2.5IN WIDTH=4IN;
ODS LISTING IMAGE_DPI=600 STYLE=JOURNAL SGE=OFF GPATH="C:\NESUG 2012";

PROC SGPLOT DATA=parmsint (WHERE=(Variable="MEDIAN_HH_ASSETS"
   AND MOD(ACCTS_PER_HH_CENTER, 0.5)=0));
TITLE "Effect of Median HH Assets on Total Revenue, 2011";
SCATTER X=ACCTS_PER_HH_CENTER Y=Estimate
   /YERRORLOWER=LCL YERRORUPPER=UCL MARKERATTRS=(SYMBOL=CIRCLE)
   ERRORBARATTRS=(PATTERN=4);
XAXIS TYPE=DISCRETE OFFSETMIN=0.02 LABEL="Accounts per HH";
YAXIS MIN=-2 MAX=4 VALUES=(-2 TO 4 BY 1)
   LABEL="b (Median HH Assets)";
REFLINE 0 /AXIS=Y TRANSPARENCY=0.5;
RUN;

ODS GRAPHICS OFF;
```

Coefficient Plot



PriceMetrix
REALIZE YOUR VALUE

Case #2

Canadian Attitudes Toward Canada–US Relations

PriceMetrix®

REALIZE YOUR VALUE

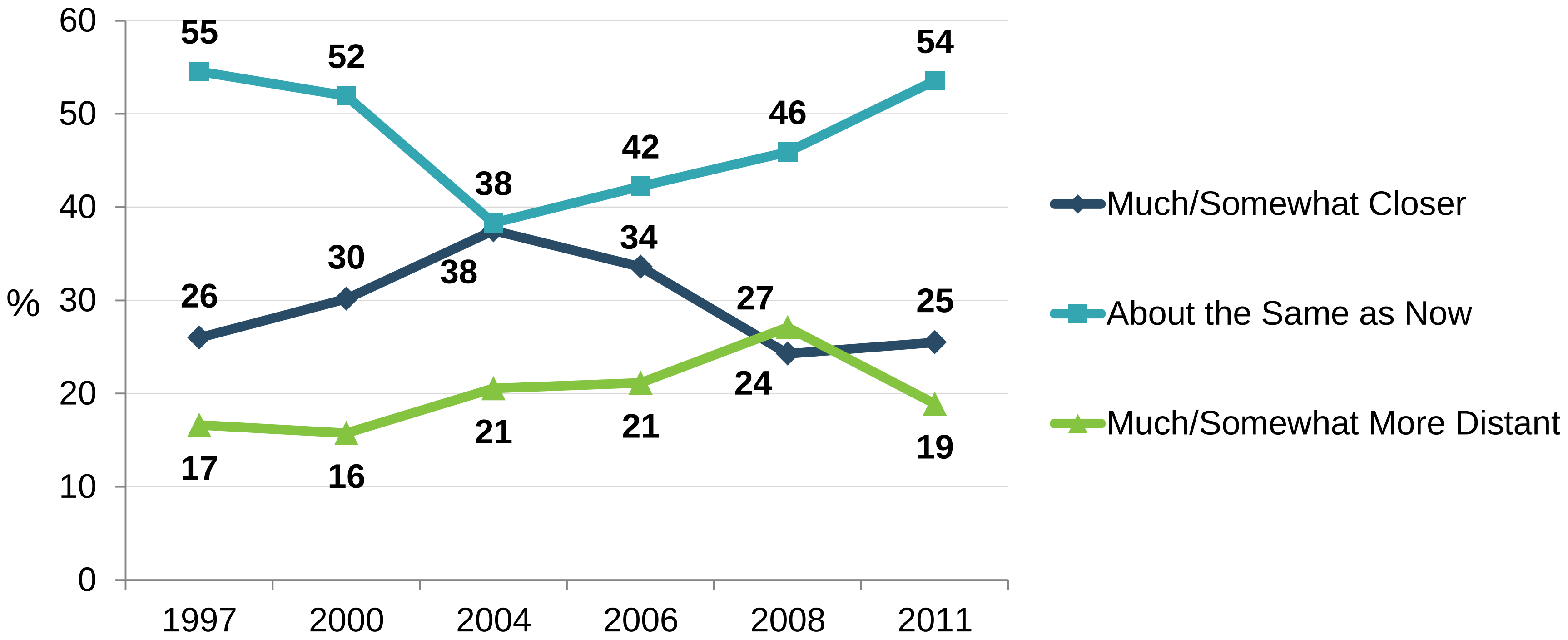
Research Questions

- What does the Canadian public think about Canada–U.S. relations?
  - What is the role of *political variables* (party identification, ideology) in shaping such attitudes?
  - What is the role of *proximity* to the US?
  - How do political variables and proximity *interact*?

PriceMetrix®
REALIZE YOUR VALUE

Data: Canadian Election Studies (1997–2011)

"Do you think Canada's ties with the United States should be much closer, somewhat closer, about the same as now, somewhat more distant, or much more distant?"



PriceMetrix
REALIZE YOUR VALUE

Let’s Interact Some More!
Specifying the Model with PROC LOGISTIC

PriceMetrix
REALIZE YOUR VALUE

Model Specification

```
PROC LOGISTIC DATA=data_7;
   MODEL CANADA_TIES_US=
   POST_PARTY_CONS POST_PARTY_NDP POST_PARTY_BQ
   POST_PARTY_OTHER POST_NO_PARTY LEFT_RIGHT
   LN_DISTANCE_USA
   /LINK=CLOGIT RSQUARE;
   WEIGHT WEIGHT;
RUN;
```

* Control variables included in the models but not shown.

PriceMetrix®
REALIZE YOUR VALUE

Model Specification

```
PROC LOGISTIC DATA=data_7 OUTEST=parmest;
   MODEL CANADA_TIES_US=
   POST_PARTY_CONS POST_PARTY_NDP POST_PARTY_BQ
   POST_PARTY_OTHER POST_NO_PARTY LEFT_RIGHT
   LN_DISTANCE_USA
   LN_DIST_USA_CONS LN_DIST_USA_NDP LN_DIST_USA_BQ
   LN_DIST_USA_OTH_PTY LN_DIST_USA_NO_PTY
   LN_DIST_USA_L_R
   /LINK=CLOGIT RSQUARE;
   INT_EFFECT1: TEST LN_DIST_USA_CONS=LN_DIST_USA_NDP=
   LN_DIST_USA_BQ=LN_DIST_USA_OTH_PTY=
   LN_DIST_USA_NO_PTY=0;
   INT_EFFECT2: TEST LN_DIST_USA_L_R=0;
   WEIGHT WEIGHT;
RUN;
```

* Control variables included in the models but not shown.

PriceMetrix
REALIZE YOUR VALUE

Model Specification

```
PROC LOGISTIC DATA=data_7 OUTEST=parmest;
   MODEL CANADA_TIES_US=
   POST_PARTY_CONS POST_PARTY_NDP POST_PARTY_BQ
   POST_PARTY_OTHER POST_NO_PARTY LEFT_RIGHT
   LN_DISTANCE_USA
   LN_DIST_USA_CONS LN_DIST_USA_NDP LN_DIST_USA_BQ
   LN_DIST_USA_OTH_PTY LN_DIST_USA_NO_PTY
   LN_DIST_USA_L_R
   /LINK=CLOGIT RSQUARE;
   INT_EFFECT1: TEST LN_DIST_USA_CONS=0,
   LN_DIST_USA_NDP=0, LN_DIST_USA_BQ=0,
   LN_DIST_USA_OTH_PTY=0, LN_DIST_USA_NO_PTY=0;
   INT_EFFECT2: TEST LN_DIST_USA_L_R=0;
   WEIGHT WEIGHT;
RUN;
```

* Control variables included in the models but not shown.

PriceMetrix

REALIZE YOUR VALUE

Results: Main Effects Model

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	39777.017	38608.978
SC	39800.129	38878.618
-2 Log L	39771.017	38538.978

R-Square 0.0724 Max-rescaled R-Square 0.0795

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1232.0390	32	<.0001
Score	1168.6213	32	<.0001
Wald	1205.9667	32	<.0001

PriceMetrix®
REALIZE YOUR VALUE

Results: Main Effects Model

Analysis of Maximum Likelihood Estimates

Parameter		Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.2800	0.0641	1264.3924	<.0001
Intercept	2	-0.9980	0.0611	266.7331	<.0001
Intercept	3	1.4221	0.0618	529.7684	<.0001
POST_PARTY_CONS		0.4089	0.0412	98.4413	<.0001
POST_PARTY_NDP		-0.5107	0.0532	92.2785	<.0001
POST_PARTY_BQ		-0.4228	0.0659	41.1882	<.0001
POST_PARTY_OTHER		-0.5187	0.1050	24.4228	<.0001
POST_NO_PARTY		0.0394	0.0431	0.8325	0.3616
LEFT_RIGHT		0.0594	0.00816	53.0024	<.0001
LN_DISTANCE_USA		-0.0253	0.0201	1.5963	0.2064

Results: Interactive Model

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	39777.017	38590.735
SC	39800.129	38906.599
-2 Log L	39771.017	38508.735

R-Square 0.0742 Max-rescaled R-Square 0.0813

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1262.2822	38	<.0001
Score	1195.3992	38	<.0001
Wald	1232.1770	38	<.0001

PriceMetrix
REALIZE YOUR VALUE

Results: Interactive Model

Analysis of Maximum Likelihood Estimates

Parameter		Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.2824	0.0642	1263.5064	<.0001
Intercept	2	-0.9999	0.0612	266.9235	<.0001
Intercept	3	1.4238	0.0619	529.4537	<.0001
POST_PARTY_CONS		0.4140	0.0413	100.5214	<.0001
POST_PARTY_NDP		-0.5161	0.0532	93.9913	<.0001
POST_PARTY_BQ		-0.4353	0.0666	42.7814	<.0001
POST_PARTY_OTHER		-0.5223	0.1052	24.6627	<.0001
POST_NO_PARTY		0.0365	0.0432	0.7123	0.3987
LEFT_RIGHT		0.0589	0.00817	51.9388	<.0001
LN_DISTANCE_USA		0.0101	0.0311	0.1055	0.7453
LN_DIST_USA_CONS		-0.1031	0.0394	6.8508	0.0089
LN_DIST_USA_NDP		0.1041	0.0518	4.0362	0.0445
LN_DIST_USA_BQ		-0.0698	0.0803	0.7541	0.3852
LN_DIST_USA_OTH_PTY		0.2790	0.1125	6.1510	0.0131
LN_DIST_USA_NO_PTY		-0.0674	0.0408	2.7301	0.0985
LN_DIST_USA_L_R		-0.00410	0.00797	0.2647	0.6069

PriceMetrix®
REALIZE YOUR VALUE

Results: Interactive Model

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
INT_EFFECT1	26.8584	5	<.0001
INT_EFFECT2	0.2647	1	0.6069

PriceMetrix®
REALIZE YOUR VALUE

Results: Summary Dataset

Partial output from the parmest dataset:

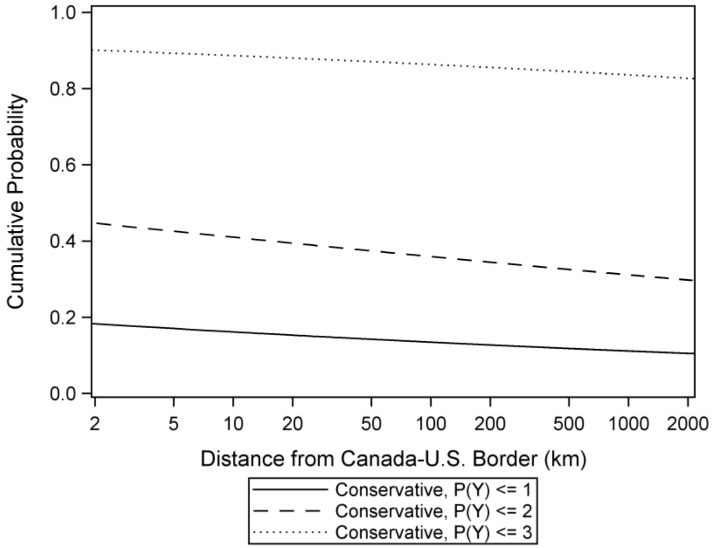
LINK	_TYPE_	_NAME_	Intercept_1	Intercept_2	Intercept_3
LOGIT	PARMS	CANADA_TIES_US	-2.28241	-0.99994	1.42381

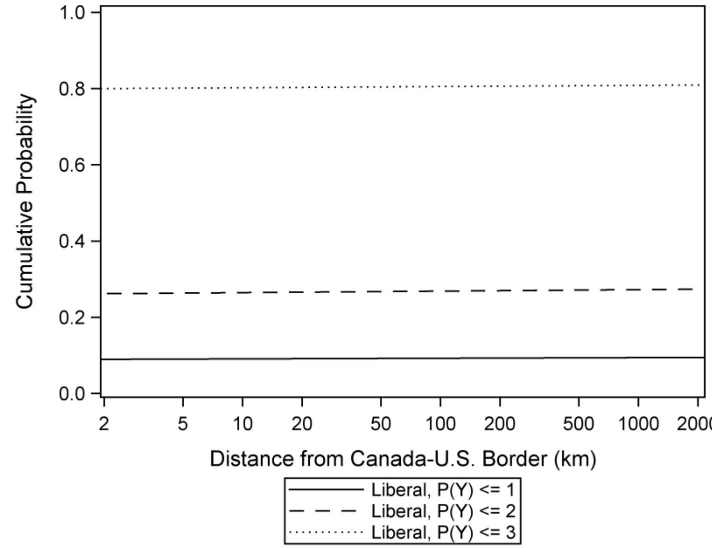
POST_PARTY_CONS	POST_PARTY_NDP	POST_PARTY_BQ	LN_DISTANCE_USA
0.41404	-0.51606	-0.43530	0.010105

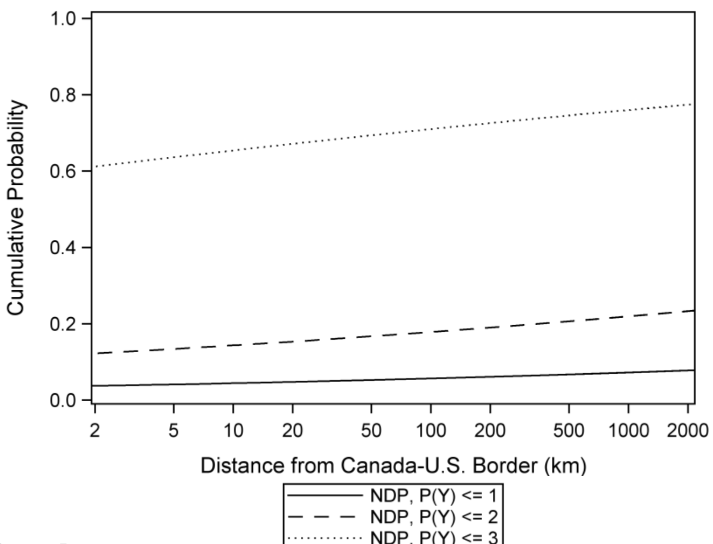
LN_DIST_USA_CONS	LN_DIST_USA_NDP	LN_DIST_USA_BQ
-0.10310	0.10415	-0.069773

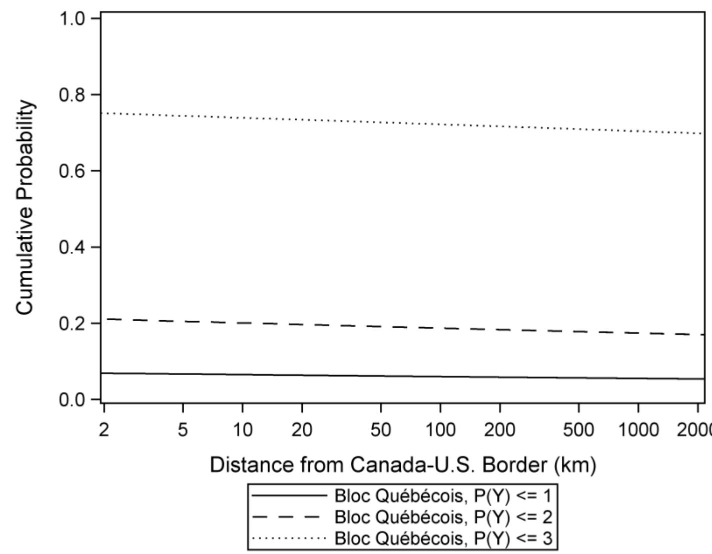
PriceMetrix®
REALIZE YOUR VALUE

Effect Plots









Effect Plot

```
DATA plot_1 (KEEP=DISTANCE_CAN_US_BORDER LN_DISTANCE_USA
   LN_DISTANCE_USA_CTR CP_: );
SET parmest (KEEP=Intercept_: POST_PARTY_: LN_DIST:
   RENAME=(POST_PARTY_CONS=b_CONS POST_PARTY_NDP=b_NDP
   POST_PARTY_BQ=b_BQ LN_DISTANCE_USA=b_LN_DISTANCE_USA
   LN_DIST_USA_CONS=b_LN_DIST_USA_CONS
   LN_DIST_USA_NDP=b_LN_DIST_USA_NDP
   LN_DIST_USA_BQ=b_LN_DIST_USA_BQ));
DO i=0.1, 0.5, 1 TO 2500;
   DISTANCE_CAN_US_BORDER=i;
   LN_DISTANCE_USA=(LOG(DISTANCE_CAN_US_BORDER));
   LN_DISTANCE_USA_CTR=(LOG(DISTANCE_CAN_US_BORDER))
      -4.6644800622;
```

PriceMetrix

REALIZE YOUR VALUE

Effect Plot

```
REG_EQN_1_LIB=Intercept_1 +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR);
CP_1_LIB=CDF('LOGISTIC',REG_EQN_1_LIB);
REG_EQN_1_CONS=Intercept_1 + (b_CONS*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_CONS*(1*LN_DISTANCE_USA_CTR));
CP_1_CONS=CDF('LOGISTIC',REG_EQN_1_CONS);
REG_EQN_1_NDP=Intercept_1 + (b_NDP*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_NDP*(1*LN_DISTANCE_USA_CTR));
CP_1_NDP=CDF('LOGISTIC',REG_EQN_1_NDP);
REG_EQN_1_BQ=Intercept_1 + (b_BQ*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_BQ*(1*LN_DISTANCE_USA_CTR));
CP_1_BQ=CDF('LOGISTIC',REG_EQN_1_BQ);
```

Effect Plot

```
REG_EQN_2_LIB=Intercept_2 +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR);
CP_2_LIB=CDF('LOGISTIC',REG_EQN_2_LIB);
REG_EQN_2_CONS=Intercept_2 + (b_CONS*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_CONS*(1*LN_DISTANCE_USA_CTR));
CP_2_CONS=CDF('LOGISTIC',REG_EQN_2_CONS);
REG_EQN_2_NDP=Intercept_2 + (b_NDP*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_NDP*(1*LN_DISTANCE_USA_CTR));
CP_2_NDP=CDF('LOGISTIC',REG_EQN_2_NDP);
REG_EQN_2_BQ=Intercept_2 + (b_BQ*1) +
    (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
    (b_LN_DIST_USA_BQ*(1*LN_DISTANCE_USA_CTR));
CP_2_BQ=CDF('LOGISTIC',REG_EQN_2_BQ);
```

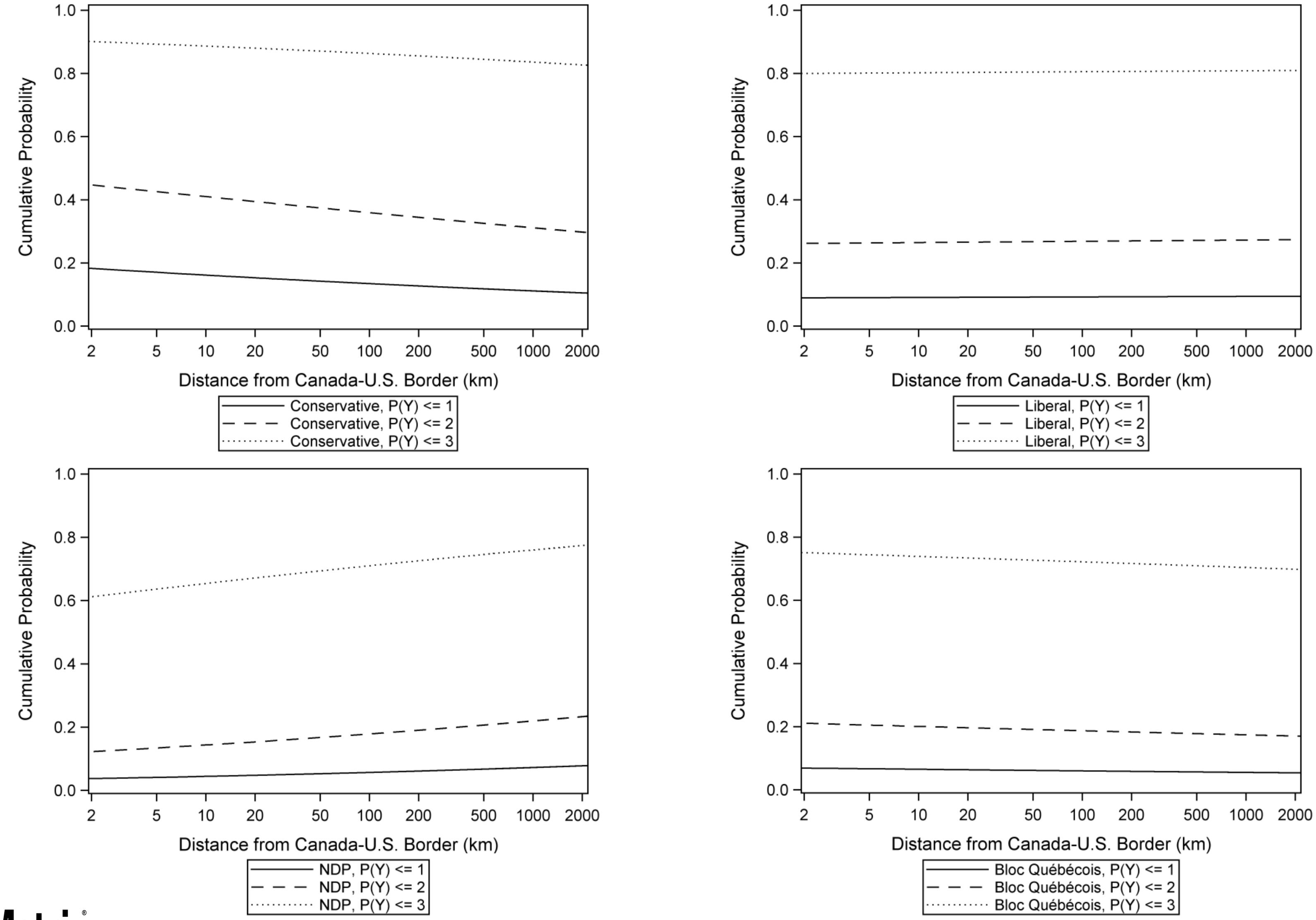
PriceMetrix®
REALIZE YOUR VALUE

Effect Plot

```
    REG_EQN_3_LIB=Intercept_3 +
        (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR);
    CP_3_LIB=CDF('LOGISTIC',REG_EQN_3_LIB);
    REG_EQN_3_CONS=Intercept_3 + (b_CONS*1) +
        (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
        (b_LN_DIST_USA_CONS*(1*LN_DISTANCE_USA_CTR));
    CP_3_CONS=CDF('LOGISTIC',REG_EQN_3_CONS);
    REG_EQN_3_NDP=Intercept_3 + (b_NDP*1) +
        (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
        (b_LN_DIST_USA_NDP*(1*LN_DISTANCE_USA_CTR));
    CP_3_NDP=CDF('LOGISTIC',REG_EQN_3_NDP);
    REG_EQN_3_BQ=Intercept_3 + (b_BQ*1) +
        (b_LN_DISTANCE_USA*LN_DISTANCE_USA_CTR) +
        (b_LN_DIST_USA_BQ*(1*LN_DISTANCE_USA_CTR));
    CP_3_BQ=CDF('LOGISTIC',REG_EQN_3_BQ);
    OUTPUT;
END;
RUN;
```

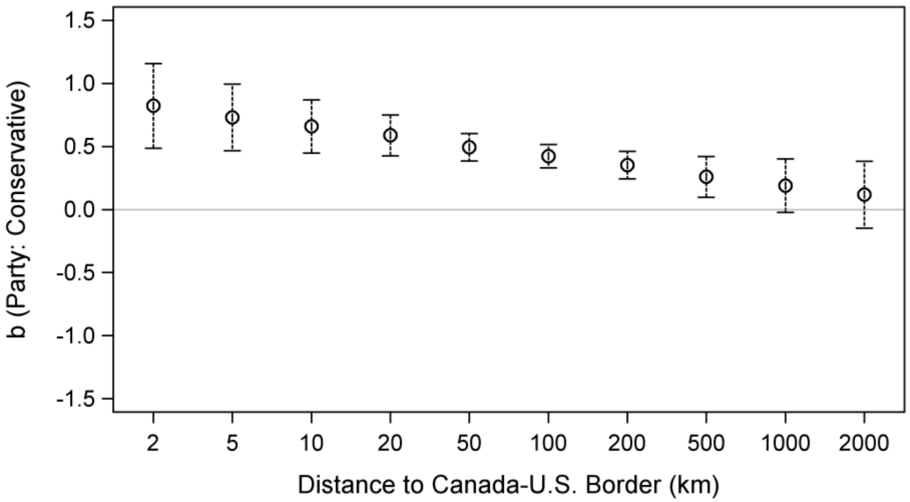
PriceMetrix®
REALIZE YOUR VALUE

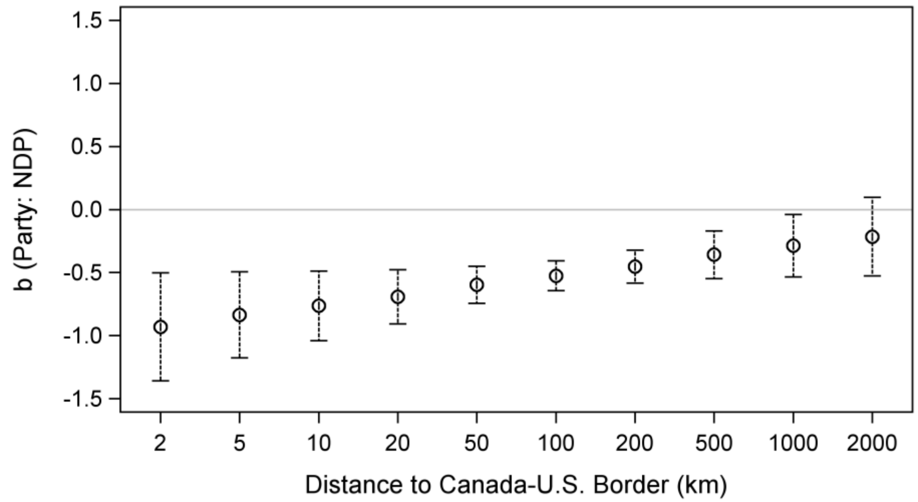
Effect Plots

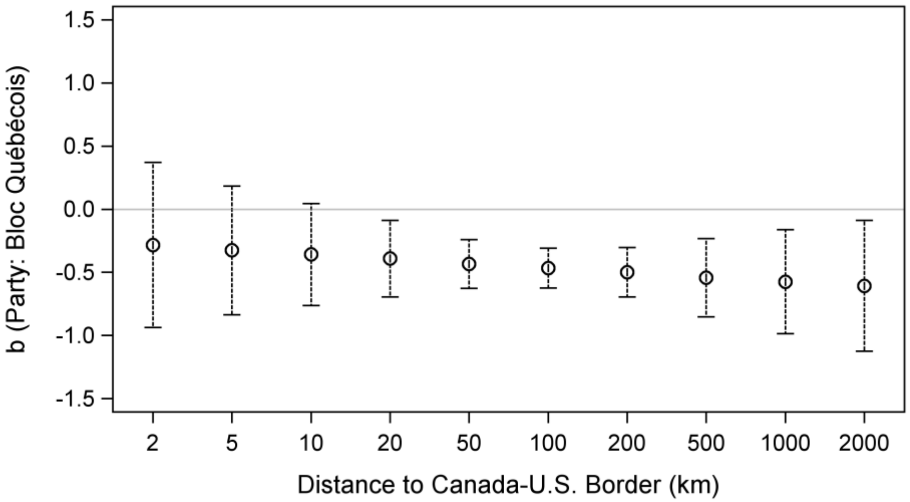


PriceMetrix®
REALIZE YOUR VALUE

Coefficient Plots







PriceMetrix
REALIZE YOUR VALUE

Recap

- Models with interaction effects require a little more care in their theorizing, specification, testing and interpretation than strictly main effects models.
- Data preparation is critical. Mean-centering is well-advised.
- Model coefficients rarely tell the entire story.
- To fully understand an interaction effect, plot it.

References

Aiken, L.S. and S.G. West (1991) *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.

Allison, P.D. (1977) "Testing for Interaction in Multiple Regression." *American Journal of Sociology* 83(1): 144–153.

Baron, R.M. and D.A. Kenny (1986) "The Moderator-Mediator Variable Distinction in Social Psychological Research." *Journal of Personality and Social Psychology* 51(6): 1173–1182.

Braumoeller, B.F. (2004) "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58(4): 807–820.

Edwards, J.R. (2008) "Seven Deadly Myths of Testing Moderation in Organizational Research." *Statistical and Methodological Myths and Urban Legends*. Eds. C.E. Lance and R.J. Vandenberg New York: Routledge.

Fox, J. (2008) *Applied Regression Analysis and Generalized Linear Models*. 2nd ed. Thousand Oaks, CA: Sage.

References

Friedrich, R.J. (1982) “In Defense of Multiplicative Terms in Multiple Regression Equations.” *American Journal of Political Science* 26(4): 797–833.

Hayes, A.F., C.J. Glynn and M.E. Huge (2012) “Cautions Regarding the Interpretation of Regression Coefficients and Hypothesis Tests in Linear Models with Interactions.” *Communication Methods and Measures* 6(1): 1–11.

Jaccard, J. and T. Dodge (2004) “Analyzing Contingent Effects in Regression Models.” *Handbook of Data Analysis*. Thousand Oaks, CA: Sage

Jaccard, J. and R. Turrisi (2003) *Interaction Effects in Multiple Regression*. 2nd ed. Thousand Oaks, CA: Sage.

Robinson, C. and R.E. Schumacker (2009) “Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues.” *Multiple Linear Regression Viewpoints* 35(1): 6–11. http://mlrv.ua.edu/2009/vol35_1/Robinson_Schumacke_rproof.pdf

Rogosa, D. (1980) “Comparing Nonparallel Regression Lines.” *Psychological Bulletin* 88(2): 307–321.

PriceMetrix®

REALIZE YOUR VALUE

Thank you!

Timothy B. Gravelle

Principal Scientist & Director, Insights Lab

tim.gravelle@pricemetrix.com