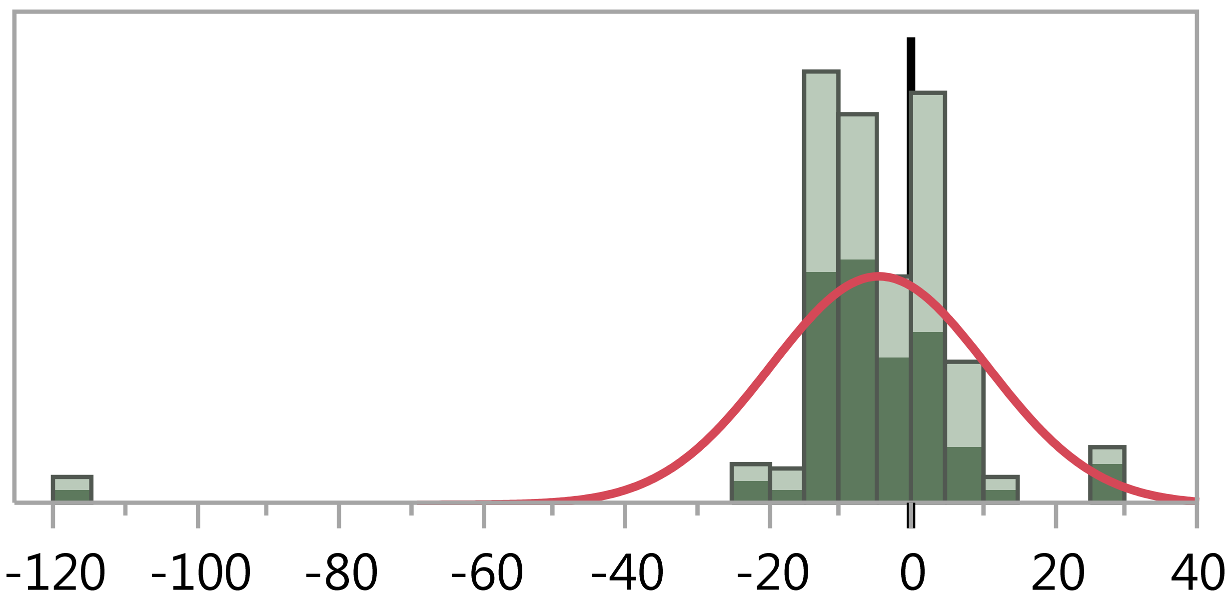
CAAA Supplemental Materials

This file provides three colored Figures plus 8 pages of Supplemental Technical Materials and Graphics supporting the paper:

Local Control Strategy: Simple Analyses of Air Pollution Data can reveal Heterogeneity in Longevity Outcomes

by Robert L. Obenchain (Risk Benefit Statistics) and S. Stanley Young (CGStat)

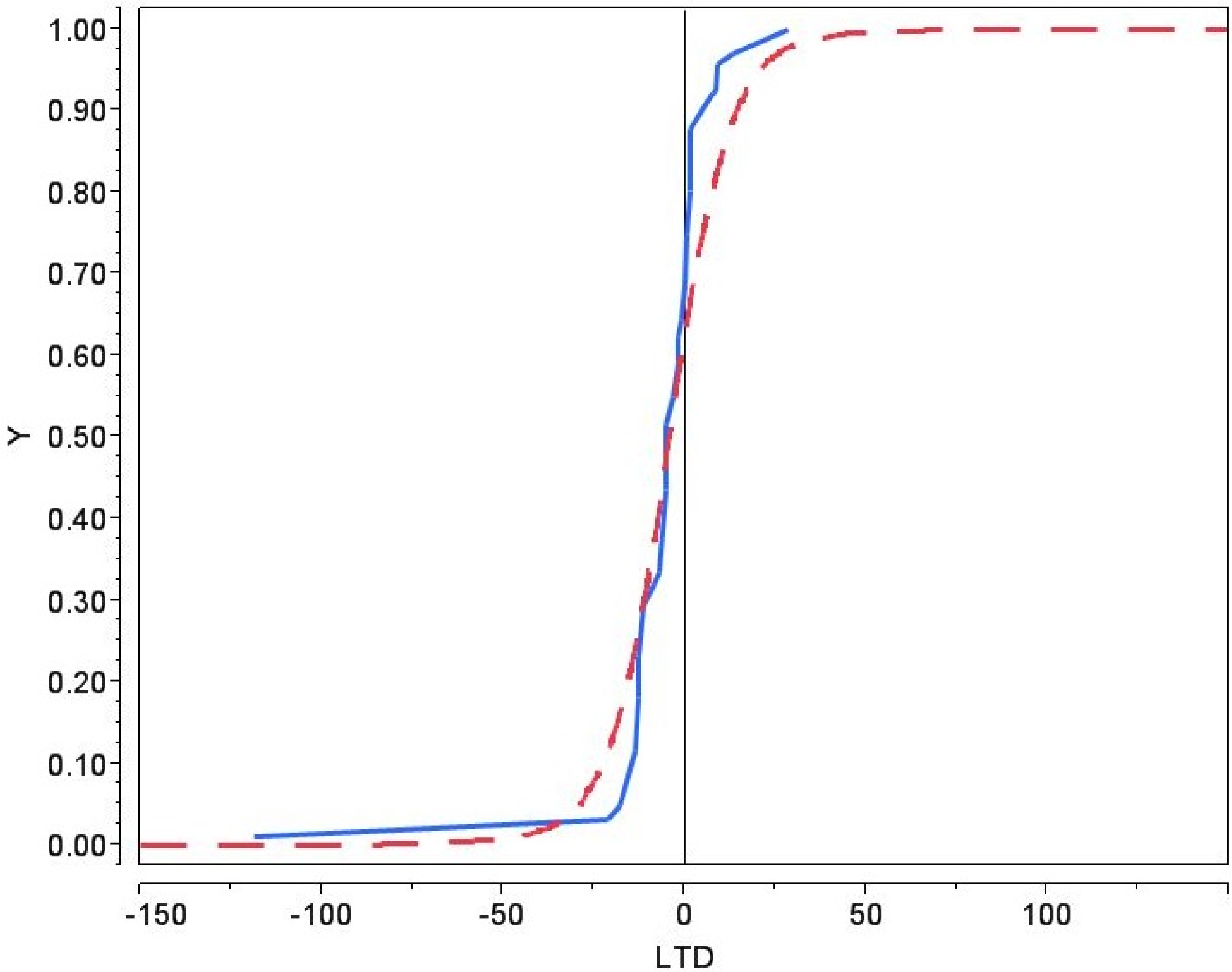
Figure 1. Distribution of Local Treatment Difference (LTD) Estimates for Change in Elderly Mortality between 1969-71 and 1972-74



Normal Fit (Mean = -4.81 Deaths per 10K Population, Std Dev = 15.17)
CAAA 1970 Compliance Classification in 1971:
Histogram Shading: Attainment Light, Nonattainment Dark

NOTE: Almost two-thirds (64.4%) of US counties have REIS characteristics that yield **negative estimates** of local differences for change in elderly mortality (LTDs) of the form [Mean Change for Nonattainment Counties minus Mean Change for Attainment Counties]. On the other hand, slightly more than one-third (35.6%) of US counties had **positive LTD estimates** ...suggesting that CAAA enforcement may have increased (rather than decreased) elderly mortality for counties with their REIS characteristics.

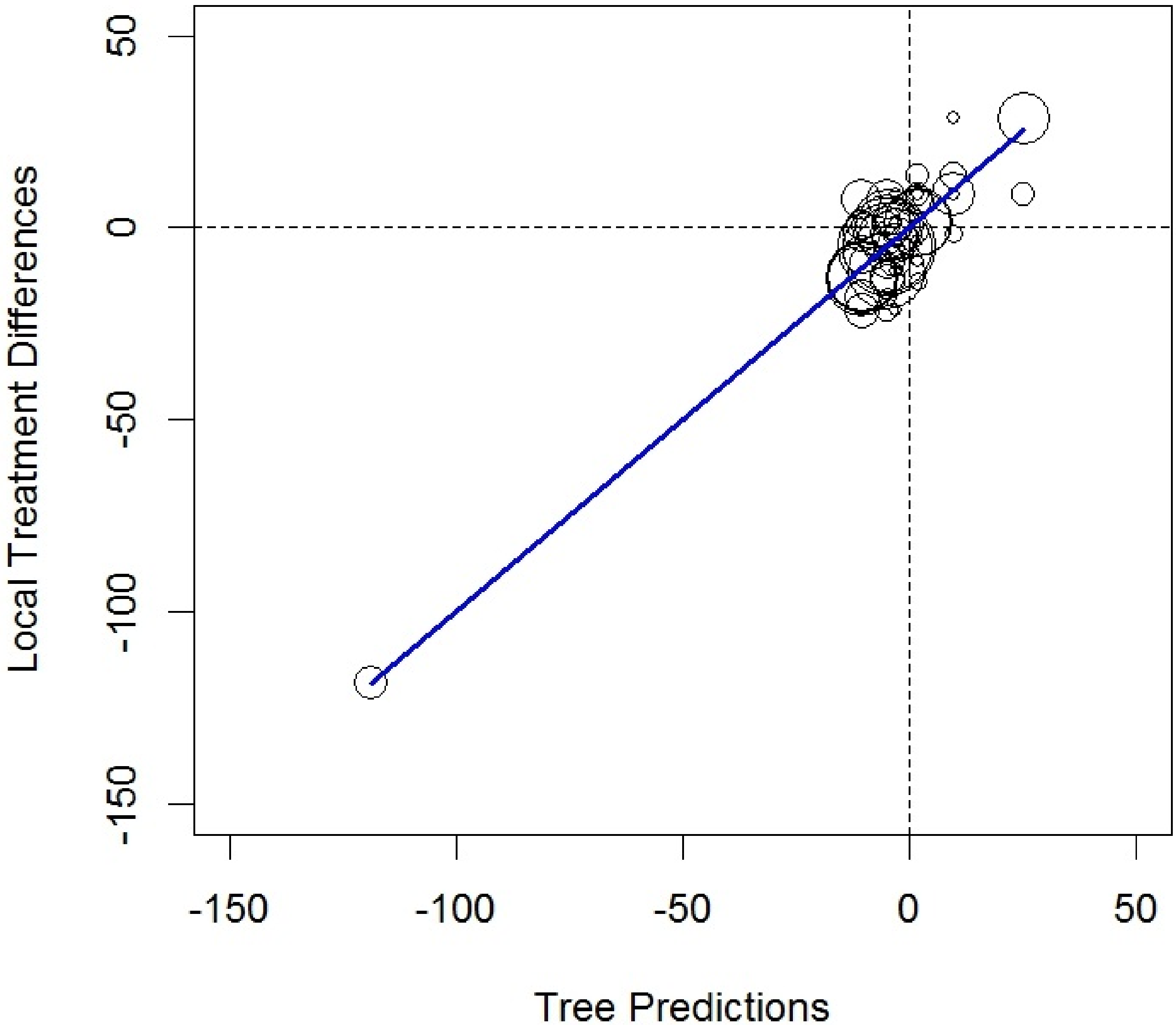
Figure 2. Cumulative Distribution Functions comparing the observed LTD distribution for Change in Elderly Mortality with its (Dashed) Random Permutation Counterpart



NOTE: The **CDF for the observed LTD distribution** for change in elderly mortality (from all 24 informative clusters using the four primary REIS ***X***-confounders) has much shorter tails than its **dashed random clustering counterpart**. In fact, the **standard deviation** of this observed LTD distribution of 15.168 elderly deaths per 10K population aged 65 to 84 is significantly less than that of its random permutation counterpart; the chi square statistic for testing that the true standard deviation is 16.451 deaths is 452.3 with a two-tailed p-value of 0.0106.

Figure 3. Goodness-of-Fit Plot:

Observed versus Predicted LTDs for Change in Elderly Mortality



Correlation = +0.928, R^2 = 0.860

NOTE: Goodness-of-fit information is quite easily *communicated visually* by simply plotting observed LTD estimates versus their predictions. Figure 3 shows such a plot displaying estimated elderly mortality LTDs for 533 US counties from 24 informative clusters against their predictions from a simple tree model with only 7 final nodes. This cross-classification of estimates and predictions yields a scatter of 53 points indicated by circles with areas proportional to the number of US counties that each point represents. The fitted solid line with slope=1 in Figure 3 also corresponds to the best fitting smoothing spline with 7 knots.

Supplemental Technical Materials and Graphics

The following materials consist primarily of **JMP Pro 11**® output for parametric **prediction** of nonparametric, observed LTDelderly **estimates** to establish their Heterogeneity. Specifically, LTDelderly effect-sizes are shown to behave like **fixed effects** ...rather than *purely random effects* from a **distribution** characterized, say, by only its mean and variance.

Each unique analysis platform within JMP software that fits statistical predictive models appears to have its own "best" way to estimate goodness-of-fit-statistics such as R-squared and adjusted R-squared.

In ordinary linear regression theory with a single X-confounder and its strictly linear generalizations (OLS), the square of the Pearson-product-moment correlation between the values of the Y-variable and their OLS predictions is the value of "adjusted" R-squared.

To make consistent, objective comparisons on goodness-of-fit of alternative statistical methodologies for predicting local treatment differences (LTDs) in elderly mortality, the Obenchain and Young paper on LC Strategy (2016) consistently uses this OLS definition of "adjusted" R-squared.

The JMP 11 Pro output documented here consists of four types:

1. Results from the JMP "Partition" platform (for the single best "tree" model) are presented first. The R-squared value reported here agrees with the OLS analysis of nested ANOVA (treatment within block) models (page 5.)

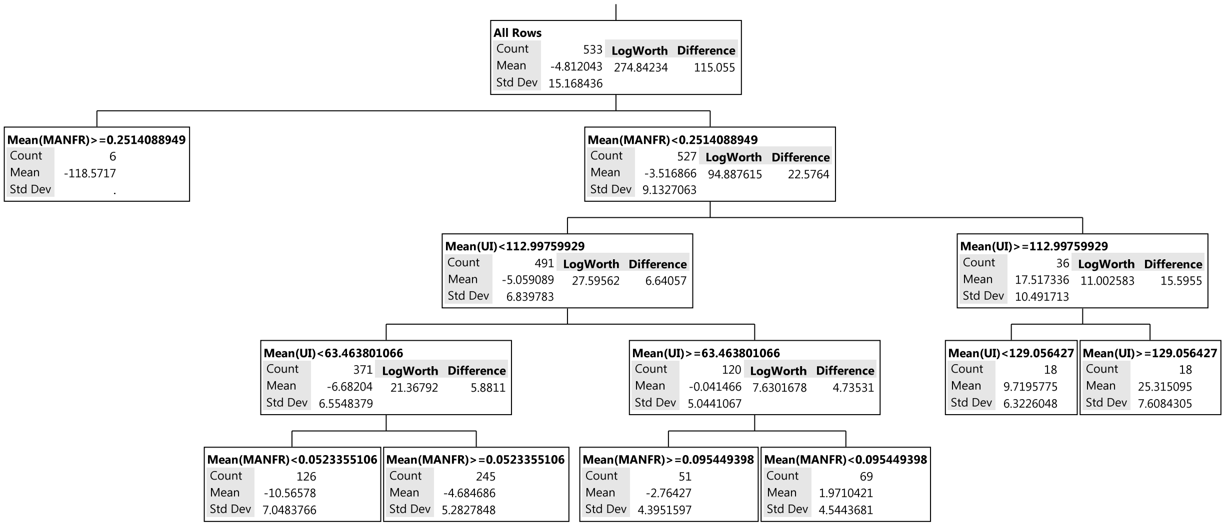
2. Results from the JMP Pro "Bootstrap Forest" platform (model averaging over 100 trees) are presented next. The R-squared values reported here not only differ slightly from the OLS definition but also would be likely to **change** slightly if the calculations were redone! Bootstrap forests are not deterministic; results depend upon the (unknown) initial random number seed value (page 6.)

3. Results from the JMP "Fit Model" platform (Multi-Variable Regression of Degree-at-most-Two) using various subsets of five ***X***-confounders (4 REIS variables plus the average TSP geometric mean) are presented third. JMP makes additional "adjustments" to R-squared for comparisons of models with different numbers of parameters / degrees-of-freedom for error. (pages 7-9.)

4. Results from the JMP "Multivariate" (correlation) platform are presented on page 10 to document the simplistic adjusted R-squared values reported in Obenchain and Young (2016). Finally, results from the JMP "Model Comparison" platform are included on page 11.

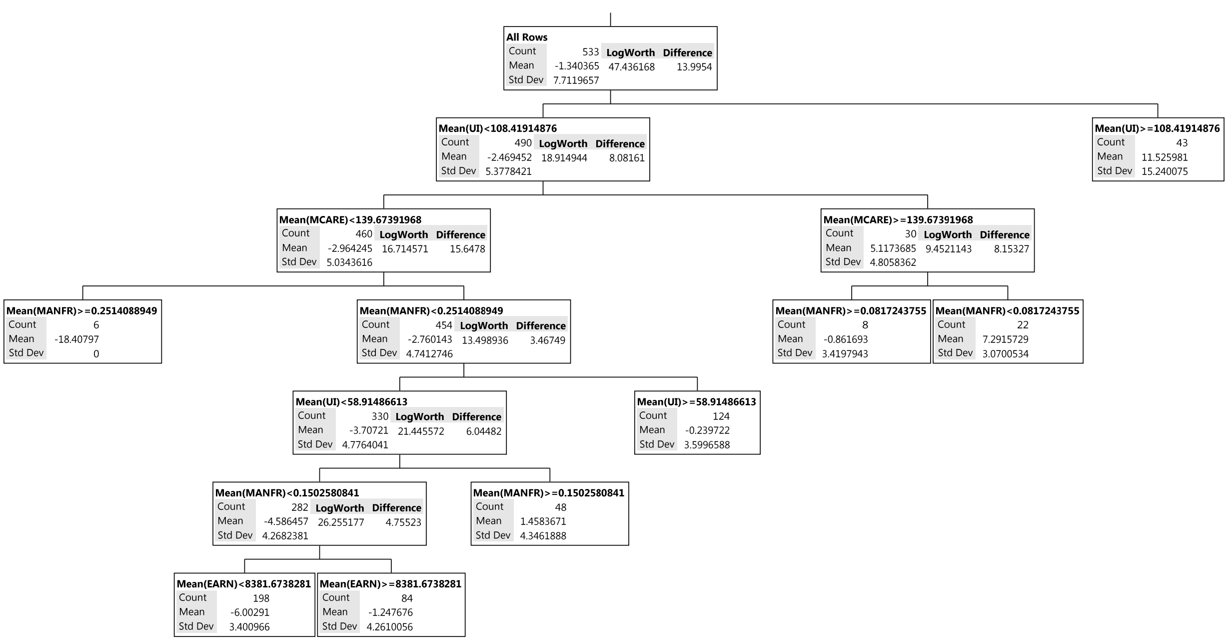
JMP Partition Tree model for LTDelderly

RSquare	RMSE	N	Number of Splits	AICc
0.861	5.656406	533	6	3376.02



JMP Partition Tree model for LTDadult

RSquare	RMSE	N	Number of Splits	AICc
0.485	5.5303807	533	7	3354.07



Bootstrap Forest predictions for LTDelderly

Specifications

Target Column:	**LTDelderly**
Number of trees in the forest:	**100**
Number of terms sampled per split:	**3**

Training rows:	533
Validation rows:	0
Test rows:	0
Number of terms:	4
Bootstrap samples:	533
Minimum Splits Per Tree:	7
Minimum Size Split:	2

Overall Statistics

Individual Trees	RMSE
In Bag	3.855238
Out of Bag	8.713975

RSquare	RMSE	N
0.900	4.791064	533

Per-Tree Summaries

Tree	Splits	Rank	OOB Loss	OOB Loss/N	RSquare	IB SSE	IB SSE/N	OOB N	OOB SSE	OOB SSE/N
1	29	99	73441.479	380.5258	0.5432	17454.312	32.747303	193	72699.062	376.67908
2	29	37	9320.8121	46.372199	0.9076	8382.5507	15.727112	201	8659.6501	43.082837
3	29	51	9486.6223	49.929591	0.9191	7320.446	13.73442	190	8376.6007	44.087372
4	29	78	21490.234	111.9283	0.8744	9951.9869	18.671645	192	20029.456	104.32008
5	29	65	10738.803	58.681984	0.8756	8305.1437	15.581883	183	9476.3407	51.783282
6	29	23	8134.0159	43.731268	0.8982	7203.179	13.514407	186	6670.2531	35.861576
7	29	48	9156.4025	49.494068	0.9144	6623.5745	12.426969	185	8269.9801	44.702595
8	29	30	9664.5122	45.587322	0.7057	18068.916	33.900406	212	9020.391	42.549014
9	29	9	7339.3555	39.672192	0.8708	8721.1966	16.36247	185	6957.4469	37.607821
10	29	80	23274.212	115.21887	0.9272	6875.9325	12.900436	202	21898.232	108.40709
...										
90	29	52	9939.5813	50.972212	0.9316	6333.5811	11.882891	195	9498.6928	48.711245
91	29	96	47171.007	237.04023	0.8676	7138.5545	13.393161	199	46930.811	235.83322
92	29	18	7892.554	42.662454	0.8770	9809.5656	18.404438	185	6669.4391	36.051022
93	29	41	8582.8223	47.158364	0.8938	8788.4376	16.488626	182	7336.3901	40.309836
94	29	85	23572.992	117.86496	0.9140	7556.2943	14.176912	200	21716.564	108.58282
95	29	73	20640.239	107.50124	0.8341	11210.779	21.033356	192	19466.908	101.39014
96	29	98	50163.263	268.25275	0.8573	7873.0284	14.77116	187	48816.442	261.05049
97	29	42	9074.4363	47.262689	0.9215	6267.7715	11.759421	192	6697.0742	34.880595
98	29	91	27452.556	150.83822	0.6259	15554.165	29.182299	182	26067.847	143.22993
99	29	74	21249.435	108.97146	0.9087	7320.3515	13.734243	195	20639.554	105.84387
100	29	64	10172.979	55.590047	0.9197	7593.807	14.247293	183	8862.7634	48.430401

Multi-Variable Regression Models for
Predicting change in elderly mortality LTDs

Primary REIS confounder **MCARE** displays no significant main-effects, squared term or interactions with the other 3 REIS variables in predicting LTDelderly.

Summary of Fit ...when using only REIS X-confounders

RSquare	0.664481
RSquare Adj	0.659358
Root Mean Square Error	8.852987
Mean of Response	-4.81204
Observations (or Sum Wgts)	533

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	8	81334.64	10166.8	129.7197
Error	524	41068.70	78.4	**Prob > F**
C. Total	532	122403.34		<.0001*

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	-9.737846	2.002457	-4.86	<.0001*
EARN	0.0004766	0.000261	1.83	0.0683
MANFR	18.180405	8.565876	2.12	0.0343*
UI	0.0613709	0.015959	3.85	0.0001*
(EARN-7811.9)*(**MANFR**-0.08772)	-0.010656	0.004368	-2.44	0.0150*
(**MANFR**-0.08772)*(UI-51.5508)	-0.838797	0.241971	-3.47	0.0006*
(EARN-7811.9)*(EARN-7811.9)	-2.247e-7	9.029e-8	-2.49	0.0131*
(**MANFR**-0.08772)*(**MANFR**-0.08772)	-1750.578	83.62746	-20.93	<.0001*
(UI-51.5508)*(UI-51.5508)	0.0024438	0.000272	8.97	<.0001*

NOTE: Due to presence of the last three quadratic terms in this (and the following) multi-variable regression model(s), predictions outside of the observed ranges of the given variables may represent potentially severe, unwarranted extrapolations.

Summary of MVReg Fit ...adding terms using MTSPGM (avg. geometric mean of TSP)

RSquare	0.675597
RSquare Adj	0.669383
Root Mean Square Error	8.721747
Mean of Response	-4.81204
Observations (or Sum Wgts)	533

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	10	82695.39	8269.54	108.7112
Error	522	39707.95	76.07	**Prob > F**
C. Total	532	122403.34		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-9.919151	2.117176	-4.69	<.0001*
Mean(EARN)	0.0003897	0.00026	1.50	0.1343
Mean(MANFR)	18.334405	8.513541	2.15	0.0317*
Mean(UI)	0.0643495	0.015768	4.08	<.0001*
Mean(MTSPGM)	**0.0095205**	**0.016039**	**0.59**	**0.5531**
(Mean(EARN)-7811.9)*(Mean(MANFR)-0.08772)	-0.013512	0.004359	-3.10	0.0020*
(Mean(MANFR)-0.08772)*(Mean(UI)-51.5508)	-0.795581	0.238631	-3.33	0.0009*
(Mean(MANFR)-0.08772)*(**Mean(MTSPGM**)-65.4098)	1.2100248	0.290336	4.17	<.0001*
(Mean(EARN)-7811.9)*(Mean(EARN)-7811.9)	-2.268e-7	8.91e-8	-2.55	0.0112*
(Mean(MANFR)-0.08772)*(Mean(MANFR)-0.08772)	-1771.899	82.81124	-21.40	<.0001*
(Mean(UI)-51.5508)*(Mean(UI)-51.5508)	0.002388	0.000269	8.89	<.0001*

Summary of Fit (without linear MTSPGM and EARN vars)

RSquare	0.673801
RSquare Adj	0.668821
Root Mean Square Error	8.729155
Mean of Response	-4.81204
Observations (or Sum Wgts)	533

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	82475.51	10309.4	135.2978
Error	524	39927.83	76.2	**Prob > F**
C. Total	532	122403.34		<.0001*

Parameter Estimates

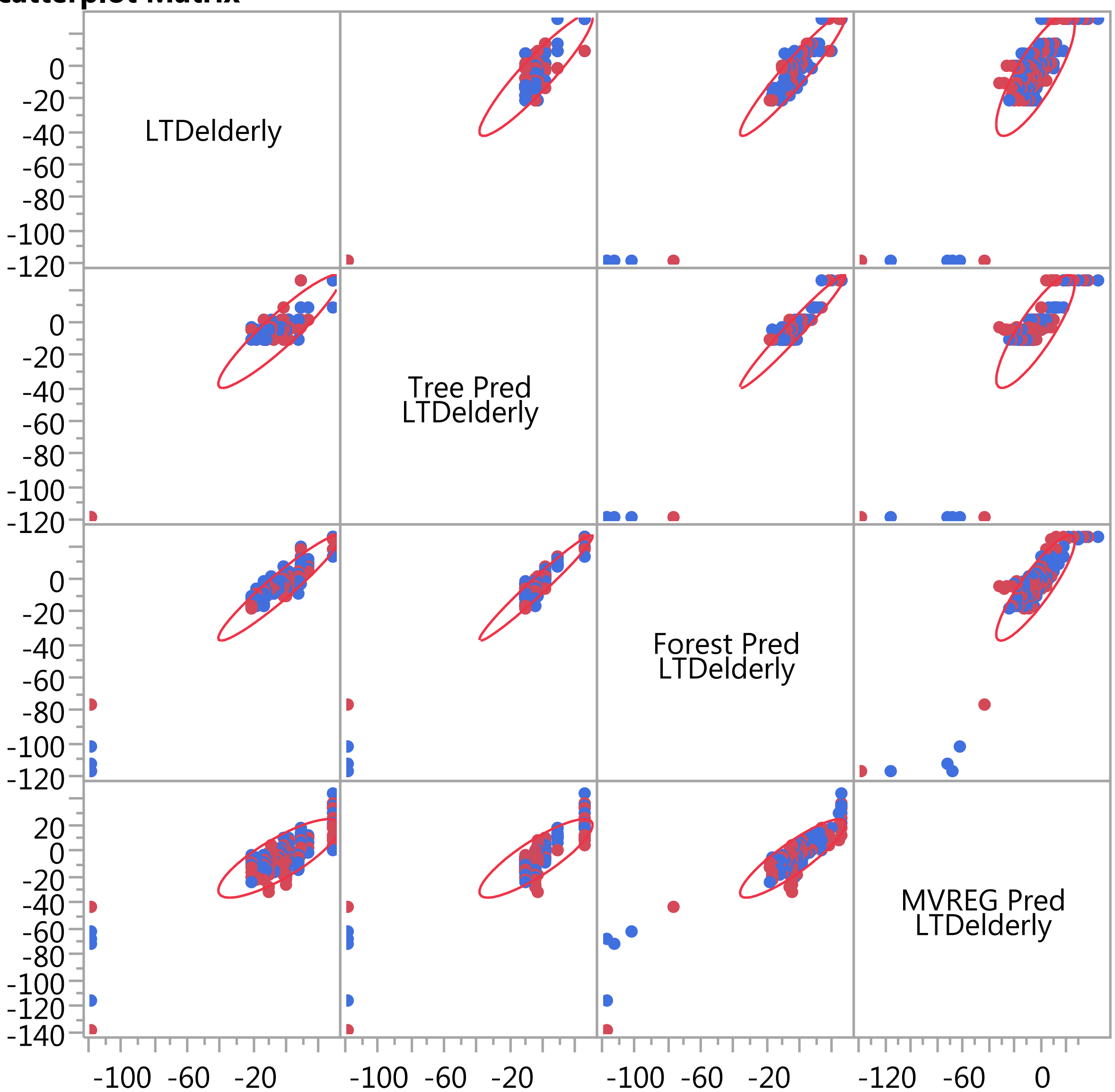
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-6.720457	0.964755	-6.97	<.0001*
Mean(**MANFR**)	22.047374	8.227374	2.68	0.0076*
Mean(UI)	0.0655815	0.015702	4.18	<.0001*
(Mean(EARN)-7811.9)*(Mean(**MANFR**)-0.08772)	-0.015458	0.004152	-3.72	0.0002*
(Mean(**MANFR**)-0.08772)*(Mean(UI)-51.5508)	-0.767888	0.238247	-3.22	0.0013*
(Mean(**MANFR**)-0.08772)*(**Mean(MTSPGM**)-65.4098)	1.2436824	0.289899	4.29	<.0001*
(Mean(EARN)-7811.9)*(Mean(EARN)-7811.9)	-1.932e-7	8.604e-8	-2.25	0.0251*
(Mean(**MANFR**)-0.08772)*(Mean(**MANFR**)-0.08772)	-1764.588	82.39129	-21.42	<.0001*
(Mean(UI)-51.5508)*(Mean(UI)-51.5508)	0.0023802	0.000269	8.85	<.0001*

The Across-6-Years Mean of the TSP Yearly Geometric Mean significantly improves prediction of change in elderly mortality LTDs only through its Interaction with MANFR when using traditional Multi-Variable Regression.

Correlations

	LTDelderly	Tree Pred LTDelderly	Forest Pred LTDelderly	MVREG Pred LTDelderly
LTDelderly	1.0000	0.9277	0.9525	0.8152
Tree Pred LTDelderly	0.9277	1.0000	0.9771	0.8521
Forest Pred LTDelderly	0.9525	0.9771	1.0000	0.8830
MVREG Pred LTDelderly	0.8152	0.8521	0.8830	1.0000

Scatterplot Matrix



attainment or nonattainment counties

Comparisons based solely upon Correlations...

LTD Predictions	Correlation with LTDs	Adjusted R-square
Single Tree Pred LTDelderly	0.9277	0.8606
Forest Pred LTDelderly	0.9525	0.9072
MVREG Pred LTDelderly	0.8152	0.6646

JMP "Model Comparisons" Output

Measures of Fit for LTDelderly

Predictor	Creator	RSquare	RASE	AAE	Freq
Tree Pred LTDelderly	Partition	0.8607	5.6564	3.9417	533
Forest Pred LTDelderly	Bootstrap Forest	0.8996	4.8018	3.3548	533
MVREG Pred LTDelderly	Fit Least Squares	0.6645	8.7779	5.7942	533

- **JMP reports the RMSE (Root Mean Squared Error) for validation and test datasets as *RASE* = Root Average Squared Error.**
- **AAE is the corresponding measure of Average Absolute Error.**