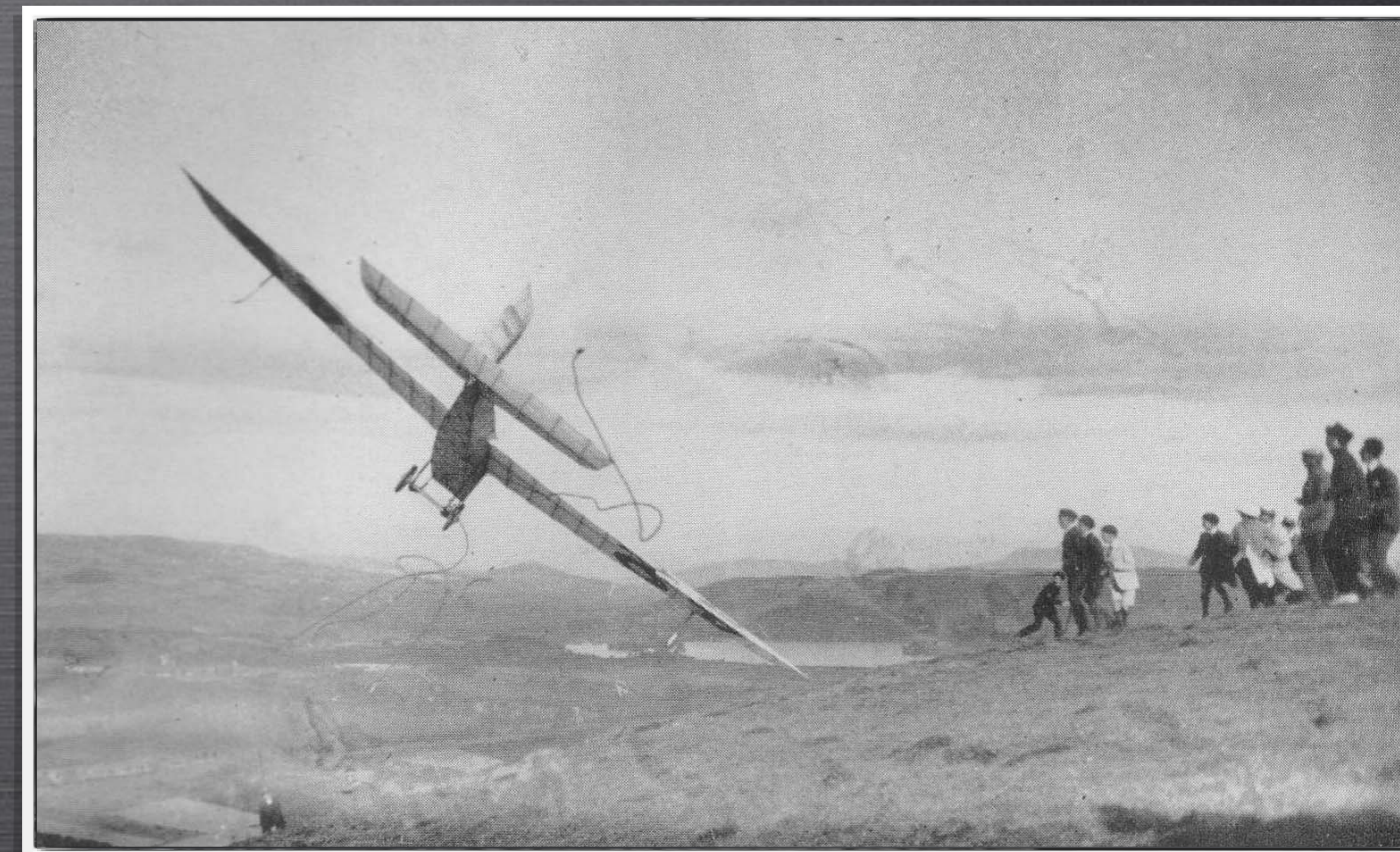
Seeing is believing: generalized added-variable plots



john gallup
Portland State University
jlgallup@pdx.edu

THE POWER OF THE SCATTERPLOT

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians
  - shows outliers and clustering

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians
  - shows outliers and clustering
- trend line summarizes the linear correlation

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians
  - shows outliers and clustering
- trend line summarizes the linear correlation
  - each observation's contribution to trend is visible

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians
  - shows outliers and clustering
- trend line summarizes the linear correlation
  - each observation's contribution to trend is visible
  - confidence interval shows closeness of fit

THE POWER OF THE SCATTERPLOT

- displays the relationship of two variables
  - intuitive to non-statisticians
  - shows outliers and clustering
- trend line summarizes the linear correlation
  - each observation's contribution to trend is visible
  - confidence interval shows closeness of fit
  - interval excluding zero slope shows significance

Scatter plot with trend line
20
10
0
-10
-20
GDP per worker growth
600
700
800
900
1000
Child survival, previous generation

ADDED-VARIABLE PLOT

ADDED-VARIABLE PLOT

- scatterplot can only show one x correlate

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates
- this is an added-variable plot

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates
- this is an added-variable plot
  - *aka* partial regression plot

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates
- this is an added-variable plot
  - *aka* partial regression plot
- displays partial correlation

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates
- this is an added-variable plot
  - *aka* partial regression plot
- displays partial correlation
  - variable *y* vs. variable of interest *x*

ADDED-VARIABLE PLOT

- scatterplot can only show one *x* correlate
- want bivariate scatter *conditional* on other correlates
- this is an added-variable plot
  - *aka* partial regression plot
- displays partial correlation
  - variable *y* vs. variable of interest *x*
  - condition on other variables X_2

CALCULATION

CALCULATION

- regress y on X_2 and take residual e_y

CALCULATION

- regress y on X_2 and take residual e_y
    - *i.e.* variation in y unaccounted for by X_2

CALCULATION

- regress y on X_2 and take residual e_y
  - *i.e.* variation in y unaccounted for by X_2
- regress x on X_2 and take residual e_x

CALCULATION

- regress y on X_2 and take residual e_y
  - *i.e.* variation in y unaccounted for by X_2
- regress x on X_2 and take residual e_x
  - *i.e.* variation in x unaccounted for by X_2

CALCULATION

- regress y on X_2 and take residual e_y
  - *i.e.* variation in y unaccounted for by X_2
- regress x on X_2 and take residual e_x
  - *i.e.* variation in x unaccounted for by X_2
- graph e_y vs. e_x

CALCULATION

- regress y on X_2 and take residual e_y
  - *i.e.* variation in y unaccounted for by X_2
- regress x on X_2 and take residual e_x
  - *i.e.* variation in x unaccounted for by X_2
- graph e_y vs. e_x
  - y and x purged of the influence of X_2

CALCULATION

- regress y on X_2 and take residual e_y
  - *i.e.* variation in y unaccounted for by X_2
- regress x on X_2 and take residual e_x
  - *i.e.* variation in x unaccounted for by X_2
- graph e_y vs. e_x
  - y and x purged of the influence of X_2
  - trend line slope *is* the OLS coefficient on x

Added-variable plot
1
0
-1
e(GDP per worker growth | X)
-45
-30
-15
0
15
30
e(Child survival, previous generation | X)

AVPLOT FEATURES

AVPLOT FEATURES

- more accessible than coefficients in a table

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient
  - shows outliers and clustering

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient
    - shows outliers and clustering
- confidence interval shows goodness of fit

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient
  - shows outliers and clustering
- confidence interval shows goodness of fit
  - statistically significant if $e_y = 0$ is excluded (for $e_x \neq 0$)

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient
  - shows outliers and clustering
- confidence interval shows goodness of fit
  - statistically significant if $e_y = 0$ is excluded (for $e_x \neq 0$)
  - zero confidence interval at $e_x = 0$

AVPLOT FEATURES

- more accessible than coefficients in a table
- see influence of each observation on estimated coefficient
  - shows outliers and clustering
- confidence interval shows goodness of fit
  - statistically significant if $e_y = 0$ is excluded (for $e_x \neq 0$)
  - zero confidence interval at $e_x = 0$
    - at this value, x is entirely accounted for by X_2

ORIGINS

ORIGINS

- developed ~1980 as visual diagnostics for outliers

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:
    - graphical representation of estimation results

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:
    - graphical representation of estimation results
    - first saw in papers by Dani Rodrick, trade economist

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:
    - graphical representation of estimation results
    - first saw in papers by Dani Rodrick, trade economist
- problem:

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:
    - graphical representation of estimation results
    - first saw in papers by Dani Rodrick, trade economist
- problem:
    - only available for OLS regression

ORIGINS

- developed ~1980 as visual diagnostics for outliers
- real potential:
  - graphical representation of estimation results
  - first saw in papers by Dani Rodrick, trade economist
- problem:
  - only available for OLS regression
  - most research uses more complex estimation

EXPANSION

EXPANSION

- oddly, no confidence intervals - **`avciplot`**

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
    - apply to fixed effects (an OLS estimation)

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- "scratch my itch" - **`xtavplot`**
  - apply to fixed effects (an OLS estimation)
  - random effects, as GLS, has OLS representation

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
  - apply to fixed effects (an OLS estimation)
  - random effects, as GLS, has OLS representation
- wider applicability: Wang (1985) for GLM

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
    - apply to fixed effects (an OLS estimation)
    - random effects, as GLS, has OLS representation
- wider applicability: Wang (1985) for GLM
- theory paper deriving avplot for

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
    - apply to fixed effects (an OLS estimation)
    - random effects, as GLS, has OLS representation
- wider applicability: Wang (1985) for GLM
- theory paper deriving avplot for
  - linear & nonlinear LS, MLE, GMM

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
    - apply to fixed effects (an OLS estimation)
    - random effects, as GLS, has OLS representation
- wider applicability: Wang (1985) for GLM
- theory paper deriving avplot for
    - linear & nonlinear LS, MLE, GMM
    - can create avplots for (almost) all estimators in Stata

EXPANSION

- oddly, no confidence intervals - **`avciplot`**
- ”scratch my itch” - **`xtavplot`**
    - apply to fixed effects (an OLS estimation)
    - random effects, as GLS, has OLS representation
- wider applicability: Wang (1985) for GLM
- theory paper deriving avplot for
  - linear & nonlinear LS, MLE, GMM
  - can create avplots for (almost) all estimators in Stata
  - will take time - must be programmed for each estimator

AV plot derivation

AV plot derivation

For the estimation equation

AV plot derivation

For the estimation equation

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

AV plot derivation

For the estimation equation

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS estimator $\mathbf{b}$ of $\boldsymbol{\beta}$:

AV plot derivation

For the estimation equation

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS estimator $\mathbf{b}$ of $\boldsymbol{\beta}$:

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

AV plot derivation

For the estimation equation

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS estimator $\mathbf{b}$ of $\boldsymbol{\beta}$:

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

Partition

AV plot derivation

For the estimation equation

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS estimator $\mathbf{b}$ of $\boldsymbol{\beta}$:

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

Partition

$$\mathbf{X} = [\mathbf{x}_1\mathbf{X}_2] \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \mathbf{b}_2 \end{bmatrix}$$

Frisch–Waugh–Lovell theorem

Frisch–Waugh–Lovell theorem

With some manipulation,

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$ "residual-maker"

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1}\mathbf{X}_2^\top$ "residual-maker"

Define $\mathbf{e}_{\mathbf{x}_1} = \mathbf{M}_2 \mathbf{x}_1$ and $\mathbf{e}_{\mathbf{y}} = \mathbf{M}_2 \mathbf{y}$.

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1}\mathbf{X}_2^\top$ "residual-maker"

Define $\mathbf{e}_{\mathbf{x}_1} = \mathbf{M}_2 \mathbf{x}_1$ and $\mathbf{e}_{\mathbf{y}} = \mathbf{M}_2 \mathbf{y}$.

Then

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$ "residual-maker"

Define $\mathbf{e}_{\mathbf{x}_1} = \mathbf{M}_2 \mathbf{x}_1$ and $\mathbf{e}_{\mathbf{y}} = \mathbf{M}_2 \mathbf{y}$.

Then

$$b_1 = (\mathbf{e}_{\mathbf{x}_1}^\top \mathbf{e}_{\mathbf{x}_1})^{-1} \mathbf{e}_{\mathbf{x}_1}^\top \mathbf{e}_{\mathbf{y}}.$$

Frisch–Waugh–Lovell theorem

With some manipulation,

$$b_1 = (\mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2^\top \mathbf{M}_2 \mathbf{y}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1}\mathbf{X}_2^\top$ "residual-maker"

Define $\mathbf{e}_{\mathbf{x}_1} = \mathbf{M}_2 \mathbf{x}_1$ and $\mathbf{e}_{\mathbf{y}} = \mathbf{M}_2 \mathbf{y}$.

Then

$$b_1 = (\mathbf{e}_{\mathbf{x}_1}^\top \mathbf{e}_{\mathbf{x}_1})^{-1} \mathbf{e}_{\mathbf{x}_1}^\top \mathbf{e}_{\mathbf{y}}.$$

Added-variable plot: graph $\mathbf{e}_{\mathbf{x}_1}$ vs. $\mathbf{e}_{\mathbf{y}}$

trend line has slope b_1.

non-OLS estimators

non-OLS estimators

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

non-OLS estimators

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}},$$

non-OLS estimators

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}},$$

then by the FWL theorem,

non-OLS estimators

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}},$$

then by the FWL theorem,

$$\hat{\beta}_1 = (\mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{x}}_1})^{-1} \mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{y}}}$$

NON-OLS ESTIMATORS

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}},$$

then by the FWL theorem,

$$\hat{\beta}_1 = (\mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{x}}_1})^{-1} \mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{y}}}$$

so we can create an added-variable plot.

NON-OLS ESTIMATORS

If we can express other estimators in terms of transformations of $\mathbf{y}$ and $\mathbf{X}$ such that

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}},$$

then by the FWL theorem,

$$\hat{\beta}_1 = (\mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{x}}_1})^{-1} \mathbf{e}_{\tilde{\mathbf{x}}_1}^\top \mathbf{e}_{\tilde{\mathbf{y}}}$$

so we can create an added-variable plot.

Essentially all estimators can be represented in the OLS form (asymptotically, everything is linear)

Generalized Least Squares

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \boldsymbol{\Sigma}$$

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \boldsymbol{\Sigma}$$

(instead of OLS's $V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \sigma^2 \mathbf{I}$)

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \boldsymbol{\Sigma}$$

$$\text{(instead of OLS's } V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \sigma^2 \mathbf{I}\text{)}$$

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y}$$

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \mathbf{\Sigma}$$

(instead of OLS's $V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \sigma^2\mathbf{I}$)

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^\top\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{\Sigma}}^{-1}\mathbf{y}$$

Transforming $\tilde{\mathbf{X}} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{X}$ and $\tilde{\mathbf{y}} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{y}$,

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \boldsymbol{\Sigma}$$

(instead of OLS's $V[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] = \sigma^2\mathbf{I}$)

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$$

Transforming $\tilde{\mathbf{X}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{X}$ and $\tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{y}$,

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}\,.$$

Generalized Least Squares

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } V[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \boldsymbol{\Sigma}$$

(instead of OLS's $V[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma^2 \mathbf{I}$)

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y}$$

Transforming $\tilde{\mathbf{X}} = \hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}$ and $\tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{y}$,

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}\,.$$

Added-variable plot of $\mathbf{e}_{\tilde{\mathbf{x}}_1}$ versus $\mathbf{e}_{\tilde{\mathbf{y}}}$.

instrumental variables (2sls)

INSTRUMENTAL VARIABLES (2SLS)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } E[\boldsymbol{\varepsilon} \mid \mathbf{X}] \neq \mathbf{0}$$
$$\text{but } E[\boldsymbol{\varepsilon} \mid \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

instrumental variables (2sls)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } E[\boldsymbol{\varepsilon} \mid \mathbf{X}] \neq \mathbf{0}$$
$$\text{but } E[\boldsymbol{\varepsilon} \mid \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

INSTRUMENTAL VARIABLES (2SLS)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } E[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] \neq \mathbf{0}$$
$$\text{but } E[\boldsymbol{\varepsilon} \,|\, \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

$$\text{Define } \hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$$

INSTRUMENTAL VARIABLES (2SLS)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \;\text{ and }\; E[\boldsymbol{\varepsilon} \,|\, \mathbf{X}] \neq \mathbf{0}$$
$$\text{but }\; E[\boldsymbol{\varepsilon} \,|\, \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

$$\text{Define } \hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$$
$$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}^\top\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^\top\mathbf{y}$$

Instrumental variables (2SLS)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } E[\boldsymbol{\varepsilon} \mid \mathbf{X}] \neq \mathbf{0}$$
$$\text{but } E[\boldsymbol{\varepsilon} \mid \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

$$\text{Define } \hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$$
$$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}^\top\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^\top\mathbf{y}$$

INSTRUMENTAL VARIABLES (2SLS)

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ and } E[\boldsymbol{\varepsilon} \mid \mathbf{X}] \neq \mathbf{0}$$
$$\text{but } E[\boldsymbol{\varepsilon} \mid \mathbf{Z}] = \mathbf{0} \text{ and } Cov[\mathbf{X}, \mathbf{Z}] \neq \mathbf{0}.$$

$$\text{Define } \hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$$
$$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}^\top\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^\top\mathbf{y}$$

Added-variable plot of $\mathbf{e}_{\hat{\mathbf{x}}_1}$ versus $\mathbf{e}_{\mathbf{y}}$.

NEXT STEPS

NEXT STEPS

- instrumental variables and system estimation

NEXT STEPS

- instrumental variables and system estimation
    - `ivregress, ivreg2, reg3`

NEXT STEPS

- instrumental variables and system estimation
  - `ivregress, ivreg2, reg3`
- more panel estimators

NEXT STEPS

- instrumental variables and system estimation
  - `ivregress, ivreg2, reg3`
- more panel estimators
  - besides `xtregress, xtregar, xtabond2`

NEXT STEPS

- instrumental variables and system estimation
  - `ivregress`, `ivreg2`, `reg3`
- more panel estimators
  - besides `xtregress`, `xtregar`, `xtabond2`
- in future

NEXT STEPS

- instrumental variables and system estimation
  - `ivregress`, `ivreg2`, `reg3`
- more panel estimators
  - besides `xtregress`, `xtregar`, `xtabond2`
- in future
  - time series: `arima`, `garch`, `var`

NEXT STEPS

- instrumental variables and system estimation
  - **`ivregress, ivreg2, reg3`**
- more panel estimators
  - besides **`xtregress, xtregar, xtabond2`**
- in future
  - time series: **`arima, garch, var`**
  - survival and limited dependent variables

NEXT STEPS

- instrumental variables and system estimation
    - **`ivregress, ivreg2, reg3`**
- more panel estimators
    - besides **`xtregress, xtregar, xtabond2`**
- in future
    - time series: **`arima, garch, var`**
    - survival and limited dependent variables
    - everything else

Blundell & Bond dynamic panel estimation

ADDITIONAL POINTS

ADDITIONAL POINTS

- added-variable plots especially good for illuminating complex estimators

ADDITIONAL POINTS

- added-variable plots especially good for illuminating complex estimators
- insightful for dummy variables

ADDITIONAL POINTS

- added-variable plots especially good for illuminating complex estimators
- insightful for dummy variables
  - whether independent or dependent

ADDITIONAL POINTS

- added-variable plots especially good for illuminating complex estimators
- insightful for dummy variables
  - whether independent or dependent
- A-V plot good for explaining estimated coefficients

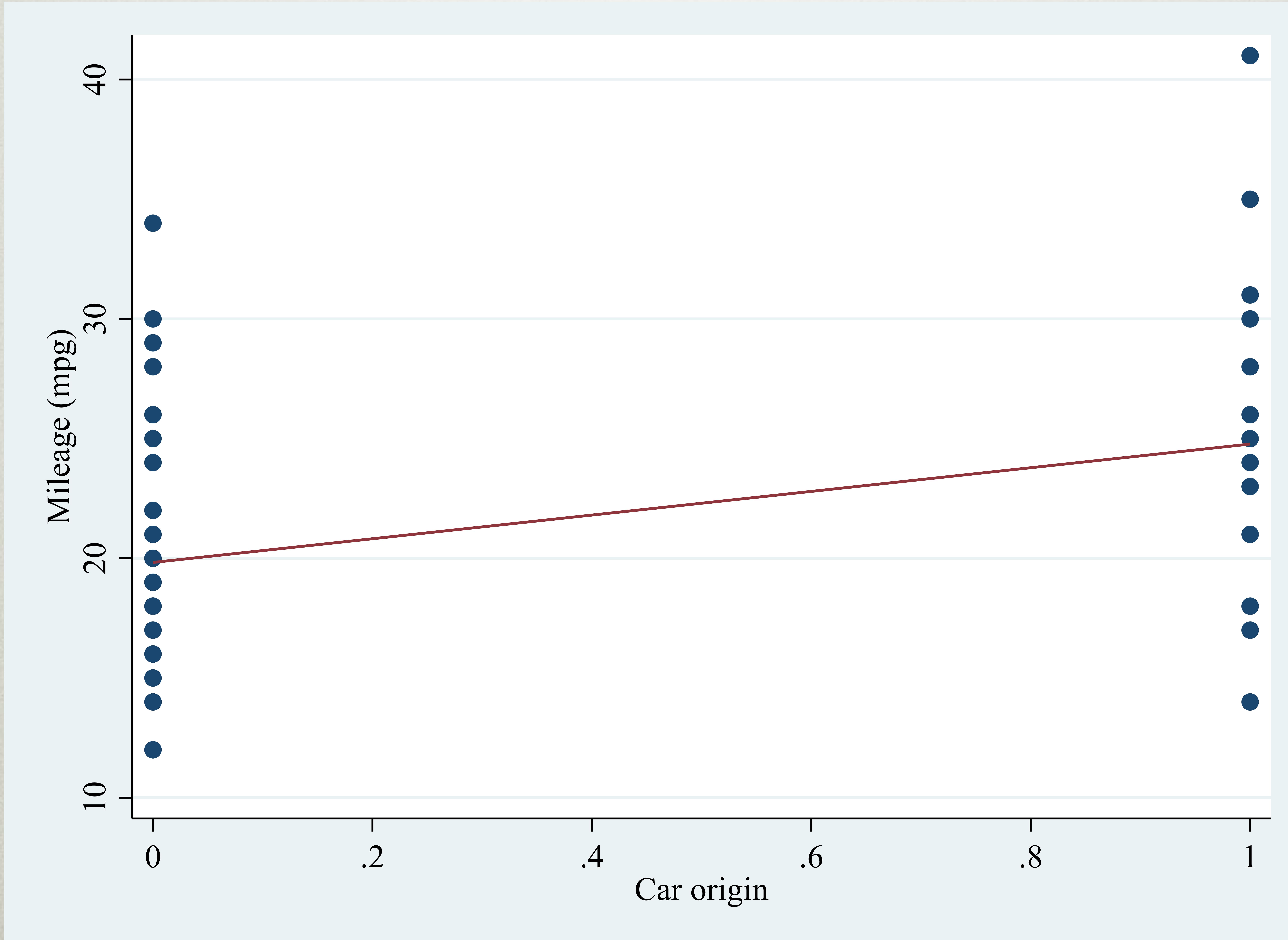
ADDITIONAL POINTS

- added-variable plots especially good for illuminating complex estimators
- insightful for dummy variables
  - whether independent or dependent
- A-V plot good for explaining estimated coefficients
  - not good for exploring functional form

ADDITIONAL POINTS

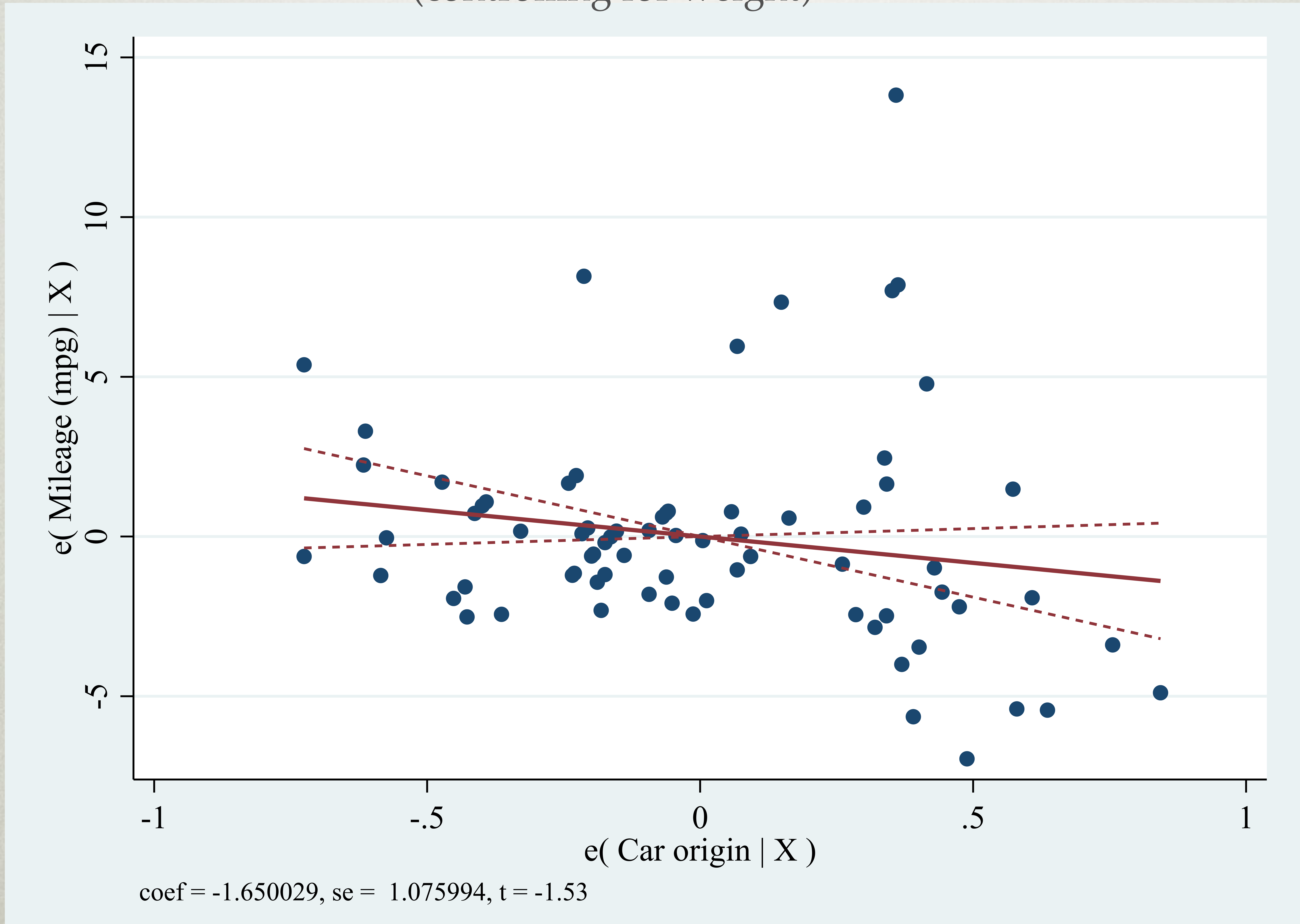
- added-variable plots especially good for illuminating complex estimators
- insightful for dummy variables
  - whether independent or dependent
- A-V plot good for explaining estimated coefficients
  - not good for exploring functional form
  - use binscatter `binsreg`, which can account for other correlates

Scatter plot of foreign dummy vs. mileage



Added-variable plot of foreign dummy vs. mileage

(controlling for weight)



NON-LINEAR ESTIMATORS

NON-LINEAR ESTIMATORS

Extremum estimators encompass MLE, NLS and GMM estimators:

NON-LINEAR ESTIMATORS

Extremum estimators encompass MLE, NLS and GMM estimators:

$$\max_{\boldsymbol{\beta}} q(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X})$$

NON-LINEAR ESTIMATORS

Extremum estimators encompass MLE, NLS and GMM estimators:

$$\max_{\boldsymbol{\beta}} q(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X})$$

Estimator $\hat{\boldsymbol{\beta}}$ implicitly defined by F.O.C.

NON-LINEAR ESTIMATORS

Extremum estimators encompass MLE, NLS and GMM estimators:

$$\max_{\boldsymbol{\beta}} q(\boldsymbol{\beta} \,|\, \mathbf{y}, \mathbf{X})$$

Estimator $\hat{\boldsymbol{\beta}}$ implicitly defined by F.O.C.

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) \equiv \frac{\partial q(\boldsymbol{\beta} \,|\, \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

where $\mathbf{H}(\boldsymbol{\beta}^*) \equiv \frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^*}$ for some $\boldsymbol{\beta}^* \in [\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0]$.

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

where $\mathbf{H}(\boldsymbol{\beta}^*) \equiv \frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^*}$ for some $\boldsymbol{\beta}^* \in [\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0]$.

Since $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$,

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

where $\mathbf{H}(\boldsymbol{\beta}^*) \equiv \frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^*}$ for some $\boldsymbol{\beta}^* \in [\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0]$.

Since $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$,

$$\boldsymbol{\beta}_0 = \mathbf{H}(\boldsymbol{\beta}^*)^{-1}\left(\mathbf{H}(\boldsymbol{\beta}^*)\hat{\boldsymbol{\beta}} + \mathbf{s}(\boldsymbol{\beta}_0)\right).$$

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

where $\mathbf{H}(\boldsymbol{\beta}^*) \equiv \frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^*}$ for some $\boldsymbol{\beta}^* \in [\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0]$.

Since $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$,

$$\boldsymbol{\beta}_0 = \mathbf{H}(\boldsymbol{\beta}^*)^{-1}\left(\mathbf{H}(\boldsymbol{\beta}^*)\hat{\boldsymbol{\beta}} + \mathbf{s}(\boldsymbol{\beta}_0)\right).$$

This is still valid when $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 = \boldsymbol{\beta}^*$, so

Take a first-order TSE of $\mathbf{s}(\hat{\boldsymbol{\beta}})$ around true parameter $\boldsymbol{\beta}_0$ and apply the Mean Value Theorem,

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{s}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta}^*)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

where $\mathbf{H}(\boldsymbol{\beta}^*) \equiv \dfrac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^*}$ for some $\boldsymbol{\beta}^* \in [\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0]$.

Since $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$,

$$\boldsymbol{\beta}_0 = \mathbf{H}(\boldsymbol{\beta}^*)^{-1}\left(\mathbf{H}(\boldsymbol{\beta}^*)\hat{\boldsymbol{\beta}} + \mathbf{s}(\boldsymbol{\beta}_0)\right).$$

This is still valid when $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 = \boldsymbol{\beta}^*$, so

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}).$$

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

if it is possible to find $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ such that

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

if it is possible to find $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ such that
$\hat{\mathbf{H}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ and $\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$.

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

if it is possible to find $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ such that
$\hat{\mathbf{H}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ and $\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$.

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

if it is possible to find $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ such that
$\hat{\mathbf{H}} = \tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}}$ and $\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}} = \tilde{\mathbf{X}}^{\top}\tilde{\mathbf{y}}$.

Then $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{y}}$

$\hat{\boldsymbol{\beta}} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}})$ allows us to linearize $\hat{\boldsymbol{\beta}}$

if it is possible to find $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ such that
$\hat{\mathbf{H}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ and $\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}} = \tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$.

Then $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$
and we can construct added-variable plots.

NONLINEAR LEAST SQUARES

NONLINEAR LEAST SQUARES

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

NONLINEAR LEAST SQUARES

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

for a known nonlinear function $\boldsymbol{\mu}(\,\cdot\,)$.

NONLINEAR LEAST SQUARES

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

for a known nonlinear function $\boldsymbol{\mu}(\,\cdot\,)$.

$$\min_{\beta} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$$

NONLINEAR LEAST SQUARES

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

for a known nonlinear function $\boldsymbol{\mu}(\,\cdot\,)$.

$$\min_{\beta} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$$

results in $\hat{\mathbf{s}} = 2\hat{\mathscr{M}}^\top(\mathbf{y} - \hat{\boldsymbol{\mu}})$ and $\hat{\mathbf{H}} = 2\hat{\mathscr{M}}^\top\hat{\mathscr{M}}$

NONLINEAR LEAST SQUARES

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

for a known nonlinear function $\boldsymbol{\mu}(\,\cdot\,)$.

$$\min_{\beta} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$$

results in $\quad \hat{\mathbf{s}} = 2\hat{\mathscr{M}}^\top(\mathbf{y} - \hat{\boldsymbol{\mu}})$ and $\hat{\mathbf{H}} = 2\hat{\mathscr{M}}^\top\hat{\mathscr{M}}$

where $\hat{\mathscr{M}} \equiv \left.\dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{y}}$$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

$$\text{where } \tilde{\mathbf{X}} = \hat{\mathscr{M}} \text{ and } \tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}\,.$$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$$\tilde{\mathbf{X}} = \hat{\mathscr{M}} \text{ and } \tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$

are adjusting for the nonlinearity of $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$

are adjusting for the nonlinearity of $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$

so that $\hat{\boldsymbol{\beta}}_{NLS}$ is a linear function of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$

$$\hat{\boldsymbol{\beta}}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$.

if $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$ were linear, $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$\tilde{\mathbf{X}} = \hat{\mathscr{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathscr{M}}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}}$

are adjusting for the nonlinearity of $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$

so that $\hat{\boldsymbol{\beta}}_{NLS}$ is a linear function of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$

and in the AV plot, $\mathbf{e}_{\tilde{\mathbf{x}}_1}$ versus $\mathbf{e}_{\tilde{\mathbf{y}}}$

$$\hat{\beta}_{NLS} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\beta} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

where $\tilde{\mathbf{X}} = \hat{\mathcal{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathcal{M}}\hat{\beta} - \hat{\mu}$.

if $\mu(\mathbf{X}, \beta)$ were linear, $\mu(\mathbf{X}, \beta) = \mathbf{X}\beta$

$\tilde{\mathbf{X}} = \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{y}$.

$\tilde{\mathbf{X}} = \hat{\mathcal{M}}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \hat{\mathcal{M}}\hat{\beta} - \hat{\mu}$

are adjusting for the nonlinearity of $\mu(\mathbf{X}, \beta)$

so that $\hat{\beta}_{NLS}$ is a linear function of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$

and in the AV plot, $\mathbf{e}_{\tilde{\mathbf{x}}_1}$ versus $\mathbf{e}_{\tilde{\mathbf{y}}}$

has a linear trend slope of $\hat{\beta}_1$.

MAXIMUM LIKELIHOOD

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

Possible for much more general MLE.

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

Possible for much more general MLE.

Let $\mathbf{y}$ be an *r.v.* in the exponential family of distributions with parameters $\boldsymbol{\theta}$ and log likelihood function

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

Possible for much more general MLE.

Let $\mathbf{y}$ be an *r.v.* in the exponential family of distributions with parameters $\boldsymbol{\theta}$ and log likelihood function

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \boldsymbol{\theta}^\top \mathbf{t}(\mathbf{y}) - a(\boldsymbol{\theta}) + b(\mathbf{y}).$$

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

Possible for much more general MLE.

Let $\mathbf{y}$ be an *r.v.* in the exponential family of distributions with parameters $\boldsymbol{\theta}$ and log likelihood function

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \boldsymbol{\theta}^{\top}\mathbf{t}(\mathbf{y}) - a(\boldsymbol{\theta}) + b(\mathbf{y}).$$

Can be shown that

$$E[\mathbf{t}(\mathbf{y})] \equiv \boldsymbol{\mu} = \frac{\partial a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad Var[\mathbf{t}(\mathbf{y})] \equiv \boldsymbol{\Sigma} = \frac{\partial^2 a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}.$$

MAXIMUM LIKELIHOOD

Derived for GLM by Wang (1985) and Pregibon (1985).

Possible for much more general MLE.

Let $\mathbf{y}$ be an *r.v.* in the exponential family of distributions with parameters $\boldsymbol{\theta}$ and log likelihood function

$$\ell(\boldsymbol{\theta} \,|\, \mathbf{y}) = \boldsymbol{\theta}^\top \mathbf{t}(\mathbf{y}) - a(\boldsymbol{\theta}) + b(\mathbf{y}).$$

Can be shown that

$$E[\mathbf{t}(\mathbf{y})] \equiv \boldsymbol{\mu} = \frac{\partial a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad Var[\mathbf{t}(\mathbf{y})] \equiv \boldsymbol{\Sigma} = \frac{\partial^2 a(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.$$

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

and $\mathscr{M} \equiv \dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}$, then

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

and $\mathscr{M} \equiv \dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}$, then

$$\hat{\mathbf{s}} = \hat{\mathscr{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{t}(\mathbf{y}) - \hat{\boldsymbol{\mu}})$$

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

and $\mathscr{M} \equiv \dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}$, then

$$\hat{\mathbf{s}} = \hat{\mathscr{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{t}(\mathbf{y}) - \hat{\boldsymbol{\mu}})$$

and $\hat{\mathbf{H}} = \hat{\mathscr{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathscr{M}}$.

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

and $\mathscr{M} \equiv \dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}$, then

$$\hat{\mathbf{s}} = \hat{\mathscr{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{t}(\mathbf{y}) - \hat{\boldsymbol{\mu}})$$

and $\hat{\mathbf{H}} = \hat{\mathscr{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathscr{M}}$.

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

and $\mathscr{M} \equiv \dfrac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\top}}$, then

$$\hat{\mathbf{s}} = \hat{\mathscr{M}}^{\top} \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{t}(\mathbf{y}) - \hat{\boldsymbol{\mu}})$$

and $\hat{\mathbf{H}} = \hat{\mathscr{M}}^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathscr{M}}$.

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\mathbf{H}}^{-1} (\hat{\mathbf{H}} \hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^{\top} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^{\top} \tilde{\mathbf{y}}$$

If the mean of $\mathbf{t}(\mathbf{y})$ is a nonlinear function of explanatory variables $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ where

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})$$

$$\text{and } \mathcal{M} \equiv \frac{\partial \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}, \text{ then}$$

$$\hat{\mathbf{s}} = \hat{\mathcal{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{t}(\mathbf{y}) - \hat{\boldsymbol{\mu}})$$

$$\text{and } \hat{\mathbf{H}} = \hat{\mathcal{M}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\mathcal{M}}.$$

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

$$\text{where } \tilde{\mathbf{X}} = \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\mathcal{M}} \text{ and } \tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}^{-1/2} \Big(\mathbf{t}(\mathbf{y}) + \hat{\mathcal{M}} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\mu}} \Big).$$

Exponential family ML provides added-variable plots for a wide range of estimators:

Exponential family ML provides added-variable plots for a wide range of estimators:

Exponential family ML provides added-variable plots for a wide range of estimators:

limited dependent variable models

Exponential family ML provides added-variable plots for a wide range of estimators:

limited dependent variable models

Exponential family ML provides added-variable plots for a wide range of estimators:

limited dependent variable models

time series (ARIMA, VAR)

Exponential family ML provides added-variable plots
for a wide range of estimators:

limited dependent variable models

time series (ARIMA, VAR)

Exponential family ML provides added-variable plots
for a wide range of estimators:

limited dependent variable models

time series (ARIMA, VAR)

survival and duration models

Exponential family ML provides added-variable plots for a wide range of estimators:

limited dependent variable models

time series (ARIMA, VAR)

survival and duration models

Exponential family ML provides added-variable plots for a wide range of estimators:

limited dependent variable models

time series (ARIMA, VAR)

survival and duration models

GENERALIZED METHOD OF MOMENTS

GENERALIZED METHOD OF MOMENTS

For elementary zero function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ and true value $\boldsymbol{\beta}_0$,

GENERALIZED METHOD OF MOMENTS

For elementary zero function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ and true value $\boldsymbol{\beta}_0$,

$$E[\mathbf{g}(\boldsymbol{\beta}_0, \mathbf{y}, \mathbf{X})] = \mathbf{0} \text{ and } E[\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})] \neq \mathbf{0} \;\forall\; \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

GENERALIZED METHOD OF MOMENTS

For elementary zero function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ and true value $\boldsymbol{\beta}_0$,

$E[\mathbf{g}(\boldsymbol{\beta}_0, \mathbf{y}, \mathbf{X})] = \mathbf{0}$ and $E[\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})] \neq \mathbf{0} \ \forall \ \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

If $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ is linear, it is $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

GENERALIZED METHOD OF MOMENTS

For elementary zero function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ and true value $\boldsymbol{\beta}_0$,

$E[\mathbf{g}(\boldsymbol{\beta}_0, \mathbf{y}, \mathbf{X})] = \mathbf{0}$ and $E[\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})] \neq \mathbf{0} \ \forall \ \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

If $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ is linear, it is $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

For possibly endogenous explanatory variables $\mathbf{X}$, let $\mathbf{Z}$ be instruments uncorrelated with $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ with $cols(\mathbf{Z}) \geq cols(\mathbf{X})$.

GENERALIZED METHOD OF MOMENTS

For elementary zero function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ and true value $\boldsymbol{\beta}_0$,

$E[\mathbf{g}(\boldsymbol{\beta}_0, \mathbf{y}, \mathbf{X})] = \mathbf{0}$ and $E[\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})] \neq \mathbf{0} \ \forall \ \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

If $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ is linear, it is $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

For possibly endogenous explanatory variables $\mathbf{X}$, let $\mathbf{Z}$ be instruments uncorrelated with $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ with $cols(\mathbf{Z}) \geq cols(\mathbf{X})$.

If $\mathbf{X}$ is uncorrelated with $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ then $\mathbf{Z} = \mathbf{X}$.

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^\top\right] = \boldsymbol{\Sigma}$.

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^{\top}\right] = \boldsymbol{\Sigma}$.

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^\top\right] = \boldsymbol{\Sigma}$.

For an assumed population moment
$E[\mathbf{Z}^\top \mathbf{g}(\boldsymbol{\beta}_\mathbf{0}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^\top\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^\top \mathbf{g}(\boldsymbol{\beta}_\mathbf{0}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^\top\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^\top \mathbf{g}(\boldsymbol{\beta}_\mathbf{0}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

$$\mathbf{Z}^\top \mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \mathbf{0} \text{ for } \hat{\boldsymbol{\beta}}_{GMM}$$

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^{\top}\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

$\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \mathbf{0}$ for $\hat{\boldsymbol{\beta}}_{GMM}$

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^{\top}\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

$$\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \mathbf{0} \text{ for } \hat{\boldsymbol{\beta}}_{GMM}$$

Accomplished by maximizing the criterion function

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^{\top}\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

$$\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \mathbf{0} \text{ for } \hat{\boldsymbol{\beta}}_{GMM}$$

Accomplished by maximizing the criterion function

$$q(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = -\frac{1}{n}\mathbf{g}(\boldsymbol{\beta})^{\top}\mathbf{Z}(\mathbf{Z}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta})$$

Let the variance of $\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ be $E\left[\mathbf{g}(\boldsymbol{\beta})\mathbf{g}(\boldsymbol{\beta})^{\top}\right] = \boldsymbol{\Sigma}$.

For an assumed population moment

$$E[\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{y}, \mathbf{X})] = \mathbf{0}$$

GMM solves the analogous sample moment

$$\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \mathbf{0} \text{ for } \hat{\boldsymbol{\beta}}_{GMM}$$

Accomplished by maximizing the criterion function

$$q(\boldsymbol{\beta} \,|\, \mathbf{y}, \mathbf{X}) = -\frac{1}{n}\mathbf{g}(\boldsymbol{\beta})^{\top}\mathbf{Z}(\mathbf{Z}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{g}(\boldsymbol{\beta})$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^{\top}\mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^{\top}\hat{\mathbf{g}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^{\top}\mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^{\top}\hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^{\top}\mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^{\top}\hat{\mathbf{G}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

where

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

where

$$\tilde{\mathbf{X}} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{Z}\hat{\mathbf{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

where

$$\tilde{\mathbf{X}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \left(\hat{\mathbf{G}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{g}}\right)$$

$$\hat{\mathbf{s}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{g}}$$

$$\hat{\mathbf{H}} = -\frac{2}{n}\hat{\mathbf{G}}^\top \mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = \hat{\mathbf{H}}^{-1}(\hat{\mathbf{H}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}) = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}$$

where

$$\tilde{\mathbf{X}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \hat{\mathbf{G}}$$

$$\tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top \left(\hat{\mathbf{G}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{g}}\right)$$

$$\hat{\boldsymbol{\Omega}} \equiv \mathbf{Z}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z} \text{ and } \hat{\mathbf{G}} \equiv \frac{\partial \mathbf{g}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.$$

All the previous estimators are GMM

Estimator	$\mathbf{g}(\boldsymbol{\beta})$	$\mathbf{Z}$	$V[\mathbf{g}]$	$\tilde{\mathbf{X}}$	$\tilde{\mathbf{y}}$
OLS	$\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$	$\mathbf{X}$	$\sigma^2\mathbf{I}$	$\mathbf{X}$	$\mathbf{y}$
GLS	$\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$	$\mathbf{X}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{X}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{y}$
2SLS	$\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$	$\mathbf{Z}$	$\sigma^2\mathbf{I}$	$\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$	$\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y}$
3SLS	$\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$	$\mathbf{Z}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}(\mathbf{Z}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}(\mathbf{Z}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y}$
MLE exp	$\mathbf{t}(\mathbf{y})-\boldsymbol{\mu}(\mathbf{X},\boldsymbol{\beta})$	$\hat{\boldsymbol{\mathcal{M}}}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\mathcal{M}}}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{t}(\mathbf{y})+\hat{\boldsymbol{\mathcal{M}}}\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\mu}})$
NLS	$\mathbf{y}-\boldsymbol{\mu}(\mathbf{X},\boldsymbol{\beta})$	$\hat{\boldsymbol{\mathcal{M}}}$	$\sigma^2\mathbf{I}$	$\hat{\boldsymbol{\mathcal{M}}}$	$\mathbf{y}+\hat{\boldsymbol{\mathcal{M}}}\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\mu}}$
NLS $\boldsymbol{\Sigma}$	$\mathbf{y}-\boldsymbol{\mu}(\mathbf{X},\boldsymbol{\beta})$	$\hat{\boldsymbol{\mathcal{M}}}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\mathcal{M}}}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{y}+\hat{\boldsymbol{\mathcal{M}}}\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\mu}})$
lin GMM	$\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$	$\mathbf{Z}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}(\mathbf{Z}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}(\mathbf{Z}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y}$
NL GMM	$\mathbf{g}(\boldsymbol{\beta},\mathbf{y},\mathbf{X})$	$\mathbf{Z}$	$\boldsymbol{\Sigma}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}(\mathbf{Z}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^\top\hat{\mathbf{G}}$	$\hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Z}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}^\top\left(\hat{\mathbf{G}}\hat{\boldsymbol{\beta}}+\hat{\mathbf{g}}\right)$

Among other GMM estimators, this adds AV plots for nonlinear instrumental variables

CONCLUDING THOUGHTS

CONCLUDING THOUGHTS

- will be implementing these AV plots in Stata

CONCLUDING THOUGHTS

- will be implementing these AV plots in Stata
- possible for Bayesian estimators?

CONCLUDING THOUGHTS

- will be implementing these AV plots in Stata
- possible for Bayesian estimators?
  - display as $\hat{\beta}_1$ surface rather than $\hat{\beta}_1$ trend line

CONCLUDING THOUGHTS

- will be implementing these AV plots in Stata
- possible for Bayesian estimators?
  - display as $\hat{\beta}_1$ surface rather than $\hat{\beta}_1$ trend line
- AV plots especially good for binary variables

CONCLUDING THOUGHTS

- will be implementing these AV plots in Stata
- possible for Bayesian estimators?
  - display as $\hat{\beta}_1$ surface rather than $\hat{\beta}_1$ trend line
- AV plots especially good for binary variables
  - dependent or independent