Modeling and Interpreting Interactions

JWAC Mini-Course TSCS

9-11 September 2009

Robert J. Franzese, Jr.

Professor of Political Science,

The University of Michigan, Ann Arbor

Overview

- Interactions in Pol-Sci: ubiquitous, but should be more
- From theory to empirical-model specification: Arguments that imply interactions (& some that don't), & how to write.
- Interpretation:
  - Effects=derivatives & differences, not coefficients!
  - Std Errs (etc.): effects vary, so do std errs (etc.)!
- Presentation: Tables & Graphs, & Choosing between equivalent Specifications
- Use & abuse of some common-practice "rules"
- Extensions:
  - Split-sample v. dummy-interaction
  - Common 2^{nd}-moment implications of interactions
  - Interactions with uncertainty = random coefficients ≈ hierarchical.

Interactions in Pol-Sci Research

- Common. '96-'01 *AJPS*, *APSR*, *JoP*:
  - 54% some stat meth (=s.e.'s), of which 24% = interax (so interax ≈ 12.5% or 1/8th total).
  - (N.b., most rest QualDep & frml thry, not counted, & "thry" in denom) so understate tech nature discipline)

Journal (1996-2001)	***Total Articles***	***Statistical Analysis***		***Interaction-Term Usage***		
		Count	*% of Tot*	*Count*	*% of Tot*	*% of Stat*
American Political Science Review	279	274	77%	69	19%	25%
American Journal Political Science	355	155	55%	47	17%	30%
Comparative Politics	130	12	9%	1	1%	8%
Comparative Political Studies	189	92	49%	23	12%	25%
International Organization	170	43	25%	9	5%	21%
International Studies Quarterly	173	70	40%	10	6%	14%
Journal of Politics	284	226	80%	55	19%	24%
Legislative Studies Quarterly	157	104	66%	19	12%	18%
World Politics	116	28	24%	6	5%	25%
TOTALS	**2446**	**1323**	**54%**	**311**	**13%**	**24%**

Interactions in Pol-Sci Theory[1]

- Ubiquitous, but our Theories/Substance say should be even more; core classes of argument inherently interactive:
  - **INSTITUTIONAL**: institutions are inherently interactive variables:
    - Institutions funnel, moderate, shape, condition, constrain, refract, magnify, augment, dampen, mitigate political processes that…
      - …translate societal interest-structures into effective political pressures,
      - …&/or pressures into public-policy responses,
      - …&/or policies to outcomes.
    - I.e., they *affect effects*≡***interaction.***

Interactions in Pol-Sci Theory^2

- Views from across institutionalist perspectives:
  - Hall: "institutionalist model=>policy more than sum countervailing pressure from soc grps; that press **mediated** by organizational dynamic."
  - Ikenberry: "[Political struggles] **mediated** by inst'l setting where [occur]"
  - Steinmo & Thelen: "inst's...**constrain & refract** politics... [effects of] macro-structures magnify or mitigated by intermediate-level inst's... help us...explain the contingent nature of pol-econ development..."
  - Shepsle: "SIE clearly a move [to] incorporating inst'l features into R-C. Structure & procedure combine w/ preferences to produce outcomes."

Interactions in Pol-Sci Theory[3]

- Ubiquitous, but our Theories/Substance say should be even more; core classes of argument inherently interactive:
  - **INSTITUTIONAL**: ...
  - **STRATEGIC**: actors' choices (outcomes) conditional upon institutional/structural environ., opportunity set, & other actors' choices.
  - **CONTEXTUAL**: actors' choices (outcomes) conditional upon environment, opportunity set, & aggregates of other actors' choices.

Interactions in Pol-Sci Theory[4]

- Across subfields:
  - Comparative Politics *examples:*
    - Electoral system & societal structure ⇒ party system.
    - Divided government & polarization ⇒ legislative productivity.
    - Corruption depends institutional & societal structures.
  - International Relations *examples:*
    - System polarity & offense-defense balance ⇒ war propensity.
    - Terrorist targeting & counterterrorism responses depend "grievance" & resources
  - American Politics …

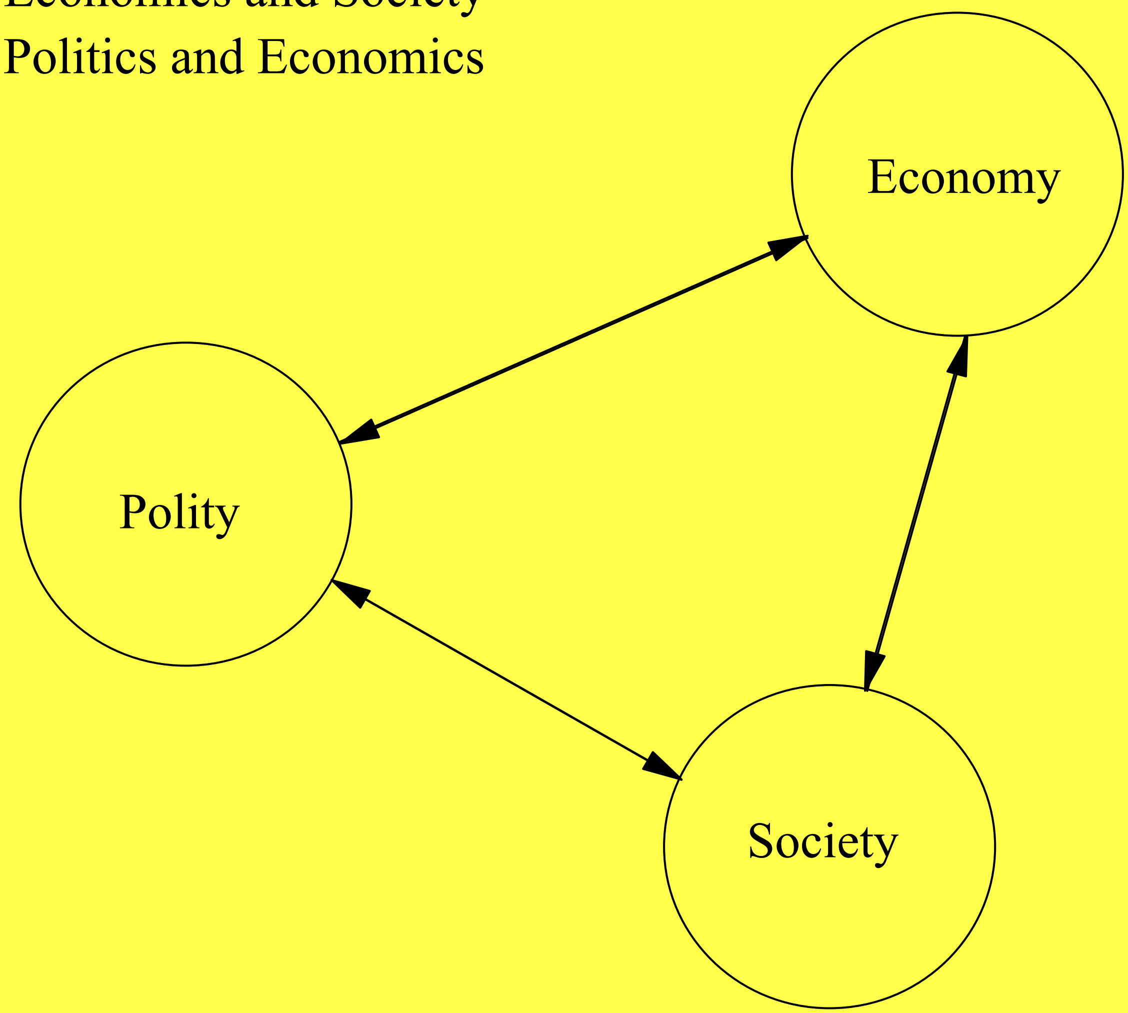
Interactions in Pol-Sci Theory[5]

- Political Economy:
  - Electoral & partisan cycles depend on inst'l & econ conditions
- Political Behavior:
  - Gov't inst's shape voter behavior: balancing (Kedar, Alesina); economic voting (Powell & Whitten); etc.
- Legislative Studies:
  - Effects divided gov't depend presidential v. parliamentary.
- Political Development:
  - Effect inequality on democratization depends cleavage structure.

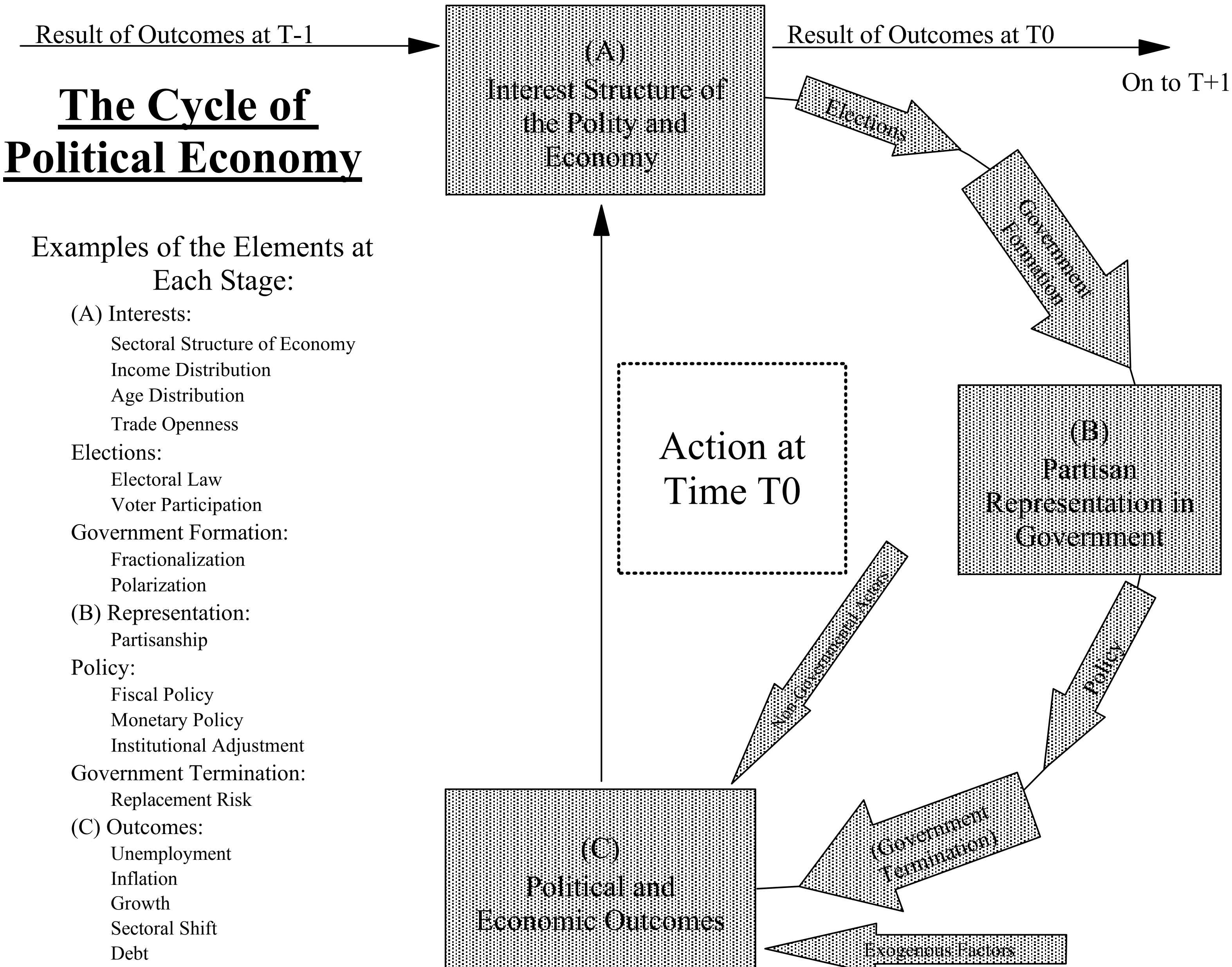
Theory & Substance:

Everyone's Favorite "Model"

Economics Affects Politics and Society
Politics Affects Economics and Society
Society Affects Politics and Economics



Theory & Substance:
An Old (& still) Favorite "Model" of Mine



The Cycle of Political Economy

Examples of the Elements at Each Stage:

- (A) Interests:
  - Sectoral Structure of Economy
  - Income Distribution
  - Age Distribution
  - Trade Openness
- Elections:
  - Electoral Law
  - Voter Participation
- Government Formation:
  - Fractionalization
  - Polarization
- (B) Representation:
  - Partisanship
- Policy:
  - Fiscal Policy
  - Monetary Policy
  - Institutional Adjustment
- Government Termination:
  - Replacement Risk
- (C) Outcomes:
  - Unemployment
  - Inflation
  - Growth
  - Sectoral Shift
  - Debt
  - Institutional Change

Theory & Substance:

An Newer Favorite "Model" of Mine

- ***Complex Context-Conditionality:***
  - Effect of (almost) everything depends on (almost) everything else.
  - E.g., Principal-Agent Situations
    - If fully principal, $y_1=f(\boldsymbol{X})$; if fully agent, $y_2=g(\boldsymbol{Z})$; institutions: $0\leq h(\boldsymbol{I})\leq 1$.

$$y = h(\mathbf{I})\,f(\mathbf{X}) + \{1 - h(\mathbf{I})\}\,g(\mathbf{Z})$$

$$\Rightarrow \frac{\partial y}{\partial x} = h(\mathbf{I})\frac{\partial f(\mathbf{X})}{\partial x} \;\; ; \;\; \frac{\partial y}{\partial z} = -h(\mathbf{I})\frac{\partial g(\mathbf{Z})}{\partial z};$$

$$\frac{\partial y}{\partial i} = \frac{\partial h(\mathbf{I})}{\partial i}\left[f(\mathbf{X}) - g(\mathbf{Z})\right]$$

(Complex) Context-Conditionality:
(Hallmark of Modern Pol-Sci Theory?)

- **Principal-Agent (Shared Control) Situations, for example:**
  - **If fully principal:** $y_1 = f(\mathbf{X})$;
  - **If fully agent:** $y_2 = g(\mathbf{Z})$;
  - **Institutions=>Monitoring & Enforcement costs principal must pay to induce agent behave as principal would:** $0 \le h(\mathbf{I}) \le 1$.
  - **RESULT:**

$$y = h(\mathbf{I}) f(\mathbf{X}) + \{1 - h(\mathbf{I})\} g(\mathbf{Z})$$

$$\Rightarrow \frac{\partial y}{\partial x} = h(\mathbf{I}) \frac{\partial f(\mathbf{X})}{\partial x} \;;$$

$$\frac{\partial y}{\partial z} = -h(\mathbf{I}) \frac{\partial g(\mathbf{Z})}{\partial z} \;;$$

$$\frac{\partial y}{\partial i} = \frac{\partial h(\mathbf{I})}{\partial i} \left[f(\mathbf{X}) - g(\mathbf{Z}) \right]$$

- **In words…**

…

…

…i.e., effect of anything depends on everything else!

Not Every Argument Is an Interactive Argument

- Not Interactive:
  - **X** affects **Y** through its effect on **Z**: **X**⇒**Z**⇒**Y**
    - In (political) psychology / behavior, this called *mediation*. Interaction is called *moderation* in this literature.
  - **X** and **Z** affect each other: **X**⇔**Z**.
    - I.e., **X** and **Z** endogenous to each other. Note: irrelevant to Gauss-Markov (OLS is BLUE); merely implies care to what partials (coefficients) mean.
  - **Y** depends on **X** controlling for **Z**, or **Y** depends on **X** & **Z**: $E(\mathbf{Y}|\mathbf{X},\mathbf{Z})=f(\mathbf{Z})$, $E(\mathbf{Y}|\mathbf{X})=f(\mathbf{Z})$, $\mathbf{Y}=f(\mathbf{X},\mathbf{Z})$
    - I.e., the outcomes differ across 2×2 of **X** and **Z**.
- <u>**Interactive**</u>: <u>***Effect***</u> ***of X on Y*** **depends on Z** (⇒ converse: Effect of **Z** on **Y** depends on **X**):

$$\frac{\partial Y}{\partial X}=f(Z)\Leftrightarrow\frac{\partial Y}{\partial Z}=f(X)$$

From Theory/Substance to Empirical-Model Specification

- **Classic Comparative-Politics Example:**
  - Societal Fragmentation, ***SFrag***, &
  - Electoral-System Proportionality, ***DMag***,
  - ⇒Effective # Parliamentary Parties: ***ENPP***
- *"Theory"*: $ENPP = f(SFrag\ ,\ DMag\ ,\ \cdot\ ,\ \varepsilon)$
- Hypotheses: $\frac{\partial ENPP}{\partial SFrag} \geq 0$ $\frac{\partial ENPP}{\partial DMag} \geq 0$

$$\frac{\partial \left\{ \frac{\partial ENPP}{\partial SFrag} \right\}}{\partial DMag} \equiv \frac{\partial \left\{ \frac{\partial ENPP}{\partial DMag} \right\}}{\partial SFrag} \equiv \frac{\partial^2 ENPP}{\partial SFrag \partial DMag} \equiv \frac{\partial^2 ENPP}{\partial DMag \partial SFrag} \geq 0$$

- Empirical Specification: Lots ways get there.

A Typical Linear-Interactive Specification

- Want linear $f(\cdot)$ w/ these properties; many ways to get there:

$$ENPP = \beta_0 + \beta_1 SFrag + \beta_2 DMag + \varepsilon$$

$$\frac{\partial ENPP}{\partial SFrag} = \beta_1 \stackrel{?}{\rightarrow} f(DMag) = \alpha_0 + \alpha_1 DMag$$

$$\frac{\partial ENPP}{\partial DMag} = \beta_2 \stackrel{?}{\rightarrow} f(SFrag) = \gamma_0 + \gamma_1 SFrag$$

$$\Rightarrow ENPP = \beta_0 + (\alpha_0 + \alpha_1 DMag)SFrag + (\gamma_0 + \gamma_1 SFrag)DMag + \varepsilon$$

$$= \beta_0 + \alpha_0 SFrag + \alpha_1 DMagSFrag + \gamma_0 DMag + \gamma_1 SFragDMag + \varepsilon$$

$$= \beta_0 + \alpha_0 SFrag + \gamma_0 DMag + (\alpha_1 + \gamma_1)SFragDMag + \varepsilon$$

$$= \beta_0 + \beta_{SF} SFrag + \beta_{DM} DMag + \beta_{SFDM} SFragDMag + \varepsilon$$

$$\Rightarrow \frac{\partial ENPP}{\partial SFrag} = \beta_{SF} + \beta_{SFDM} DMag$$

$$\frac{\partial ENPP}{\partial DMag} = \beta_{DM} + \beta_{SFDM} SFrag$$

Interpretation of *Effects*:

Derivatives & Differences, *Not* Coefficients

- Standard Linear Interactive Model:

$$EN = \beta_0 + \beta_{SF} SF + \beta_{DM} DM + \beta_{SFDM} SF \times DM + \ldots + \varepsilon$$

- Effect of $\boldsymbol{SFrag}$ on $\boldsymbol{ENPP}$ (is a *function* of $\boldsymbol{DMag}$):

$$Effect(SF) \equiv \frac{\partial EN}{\partial SF} = \beta_{SF} + \beta_{SFDM} DM$$

$$\Delta EN = \beta_{SF} \Delta SF + \beta_{SFDM} DM \cdot \Delta SF$$

$$\equiv \frac{\Delta EN}{\Delta SF} = \beta_{SF} + \beta_{SFDM} DM$$

- Effect of $\boldsymbol{DMag}$ on $\boldsymbol{ENPP}$ (is f of $\boldsymbol{SFrag}$):

$$Effect(DMag) \equiv \frac{\partial ENPP}{\partial DMag} = \beta_{DM} + \beta_{SFDM} SFrag$$

$$\equiv \Delta ENPP = \beta_{DM} \Delta DM + \beta_{SFDM} SFrag \cdot \Delta DM$$

$$\equiv \frac{\Delta ENPP}{\Delta DM} = \beta_{DM} + \beta_{SFDM} SFrag$$

Interpretation of *Effects*: NOTES[1]

- "Main Effect" & "Interactive Effect":
  - For example, β_{SF} = "*main effect* of **SFrag**"
  - **....but** β_{SF} is merely the effect of **SFrag** at other variable(s) involved in interaction with it=0, so:
    - *Other-var(s)=0* may be extreme in the sample, or beyond sample range, or even logically impossible.
    - *Other-var(s)=0* substantive meaning of 0 altered by rescaling
      - E.g., by "centering" (centering changes nothing, btw…)
    - *Other-var(s)=0* may not have anything substantively *main* about it
  - Is no Main Effect or separately & Interactive Effect; is just the effect, which conditional, varies:

$$Effect\left(SF\right) \equiv \frac{\partial EN}{\partial SF} = \beta_{SF} + \beta_{SFDM} DM \quad ; \quad Effect\left(DM\right) \equiv \frac{\partial EN}{\partial DM} = \beta_{DM} + \beta_{SFDM} SF$$

Interpretation of *Effects*: NOTES²

- **COEFFICIENTS ARE NOT EFFECTS. EFFECTS ARE DERIVATIVES &/OR DIFFERENCES.**
  - Only in ***purely*** linear-additive-separable model are they equal because only there do derivatives simply = coefficients.
  - β_{SF} is ***not*** "effect of **SFrag** 'independent of'…" & definitely not its "effect 'controlling for'…other variable(s) in the interaction"
- Cannot substitute linguistic invention for under-standing model's logic (its simple math)

Interpretation of *Effects*: NOTES³

- Interactions are logically symmetric:
  - For any function, not just lin-add.
  - If argue effect x depends z, must also believe effect z depends x.

$$\frac{\partial\left\{\frac{\partial y}{\partial x}\right\}}{\partial z} \equiv \frac{\partial\left\{\frac{\partial y}{\partial z}\right\}}{\partial x} \equiv \frac{\partial^2 y}{\partial x \partial z} \equiv \frac{\partial^2 y}{\partial z \partial x}$$

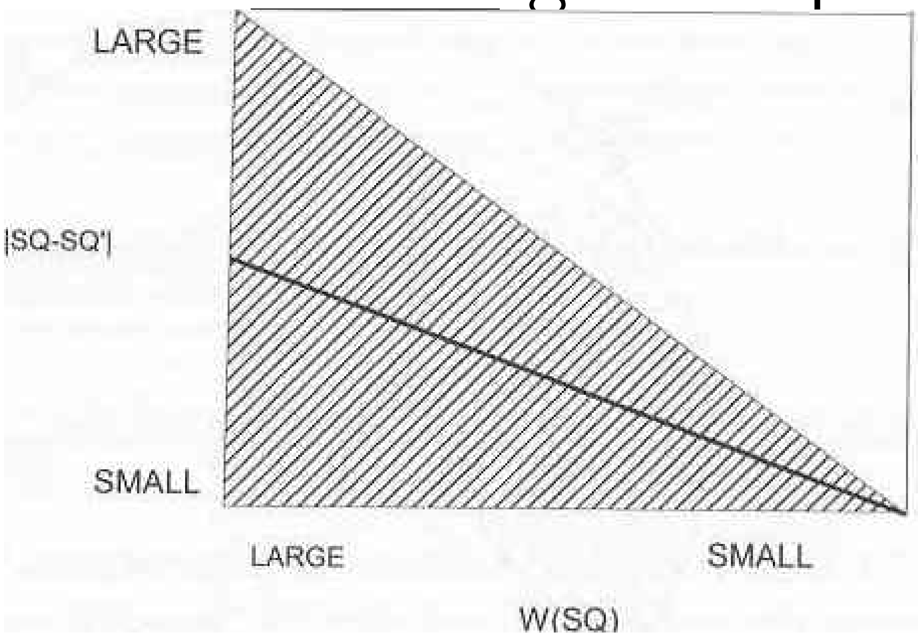
- Interactions often have 2nd-moment (variance, i.e., heteroskedacity) implications too:
  - **Larger district magnitudes, *DMag***, are "permissive" elect sys: *allow* more parties…
  - Fewer ***Veto Actors*** *allow* greater policy-change… (both need additional assumpts)
- All of this holds for any type of variable:
  - Measurement: binary, continuous…
  - ***Level***: micro or macro; i, j, k, …

Frequent 2nd-Moment Implications Interactions

- ***DMag*** permissive ele sys: ***allows*** more parties…

  $NP = \beta_0 + \beta_1 DM + \varepsilon$; $V(\varepsilon) = f(DM)$, e.g., $\sigma_0 + \sigma_1 DM$

  – Note: unmodeled interactions look like heteroskedasticity; that's general, actually. Anything unmodeled gets into $\mathbf{e}^2$…
- Few ***Veto Actors allows*** greater policy-change…



$y = \beta_0 + \beta_1 VP + \varepsilon$; $V(\varepsilon) = f(VP)$, e.g., $\sigma_0 + \sigma_1 VP$

- I.e., these are Rndm-Coeff &/or Het-sked Props…

Interpretation of *Effects*:

Standard Errors for *Effects*

$$ENPP = \beta_0 + \beta_{SF} SFrag + \beta_{DM} DMag + \beta_{SFDM} SFragDMag + \ldots + \varepsilon$$

- Std Errs reported with regression output are for coefficients, not for effects.
  - The s.e. (t-stat, p-level) for $\hat{\beta}_{SF}$ is std. err. for est'd effect ***SFrag*** at ***DMag***=0 (…which is logically impossible).
- Effect of x depends on z & v.v. (i.e., which was the point, remember?), so does the s.e.:

$$Effect(x) \equiv \frac{\partial y}{\partial x} = \beta_x + \beta_{xz} z \;\Rightarrow\; Est.Eff.(x) \equiv E\left(\frac{\partial y}{\partial x}\right) = \hat{\beta}_x + \hat{\beta}_{xz} z$$

$$Est.Var.\{Est.Eff.(x)\} \equiv E\left[Var\left\{E\left(\frac{\partial y}{\partial x}\right)\right\}\right] = E\left[Var\left\{\hat{\beta}_x + \hat{\beta}_{xz} z\right\}\right]$$

$$= \widehat{V\left\{\hat{\beta}_x + \hat{\beta}_{xz} z\right\}} = \widehat{V\left\{\hat{\beta}_x\right\}} + \widehat{V\left\{\hat{\beta}_{xz}\right\}} \cdot z^2 + 2 \cdot \widehat{C\left(\hat{\beta}_x, \hat{\beta}_{xz}\right)} z$$

- In words… More Generally: $\widehat{V(\mathbf{x}'\hat{\boldsymbol{\beta}})} = \mathbf{x}'\left[\widehat{V(\hat{\boldsymbol{\beta}})}\right]\mathbf{x}$

From Hypotheses to Hypotheses Tests:

Does Y Depend on X or Z?

$$ENPP = \beta_0 + \beta_{SF} SFrag + \beta_{DM} DMag + \beta_{SFDM} SFragDMag + \ldots + \varepsilon$$

Hypothesis	*Mathematical Expression*[91]	*Statistical test*
x affects y, or *y is a function of (depends on) x*	$y=f(x)$ $\partial y/\partial x = \beta_x + \beta_{xz} z \neq 0$	*F- test:* H_0: $\beta_x = \beta_{xz} = 0$
x increases y	$\partial y/\partial x = \beta_x + \beta_{xz} z > 0$	*Multiple t-tests:* H_0: $\beta_x + \beta_{xz} z \leq 0$
x decreases y	$\partial y/\partial x = \beta_x + \beta_{xz} z < 0$	*Multiple t- tests:* $\beta_x + \beta_{xz} z \geq 0$
z affects y, or *y is a function of (depends on) z*	$y=g(z)$ $\partial y/\partial z = \beta_z + \beta_{xz} x \neq 0$	*F- test:* H_0: $\beta_z = \beta_{xz} = 0$
z increases y	$\partial y/\partial z = \beta_z + \beta_{xz} x > 0$	*Multiple t-tests:* H_0: $\beta_z + \beta_{xz} x \leq 0$
z decreases y	$\partial y/\partial z = \beta_z + \beta_{xz} x < 0$	*Multiple t- tests:* H_0: $\beta_z + \beta_{xz} x \geq 0$

From Hypotheses to Hypotheses Tests:

Is Y's Dependence on X Conditional on Z & v.v.? How?

Hypothesis	*Mathematical Expression*[92]	*Statistical test*
The effect of x on y depends on z	$y=f(xz,\bullet)$ $\partial y/\partial x=\beta_x+\beta_{xz}z=g(z)$ $\partial(\partial y/\partial x)/\partial z=\partial^2 y/\partial x\partial z=\beta_{xz}=0$	*t-test:* H_0: $\beta_{xz}=0$
The effect of x on y increases in z	$\partial(\partial y/\partial x)/\partial z=\partial^2 y/\partial x\partial z=\beta_{xz}>0$	*t-test:* H_0: $\beta_{xz}\leq 0$
The effect of x on y decreases in z	$\partial(\partial y/\partial x)/\partial z=\partial^2 y/\partial x\partial z=\beta_{xz}<0$	*t-test:* H_0: $\beta_{xz}\geq 0$
The effect of z on y depends on x	$y=f(xz,\bullet)$ $\partial y/\partial z=\beta_z+\beta_{xz}x=h(x)$ $\partial(\partial y/\partial z)/\partial x=\partial^2 y/\partial z\partial x=\beta_{xz}=0$	*t-test:* H_0: $\beta_{xz}=0$
The effect of z on y increases in x	$\partial(\partial y/\partial z)/\partial x=\partial^2 y/\partial z\partial x=\beta_{xz}>0$	*t-test:* H_0: $\beta_{xz}\leq 0$
The effect of z on y decreases in x	$\partial(\partial y/\partial z)/\partial x=\partial^2 y/\partial z\partial x=\beta_{xz}<0$	*t-test:* H_0: $\beta_{xz}\geq 0$

Does Y Depend on X, Z, or XZ?

Hypothesis	*Mathematical Expression*[93]	*Statistical Test*
y is a function of (depends on) z, z, and/or their interaction	$y=f(x,z,xz)$	*F-test:* H_0: $\beta_x=\beta_z=\beta_{xz}=0$

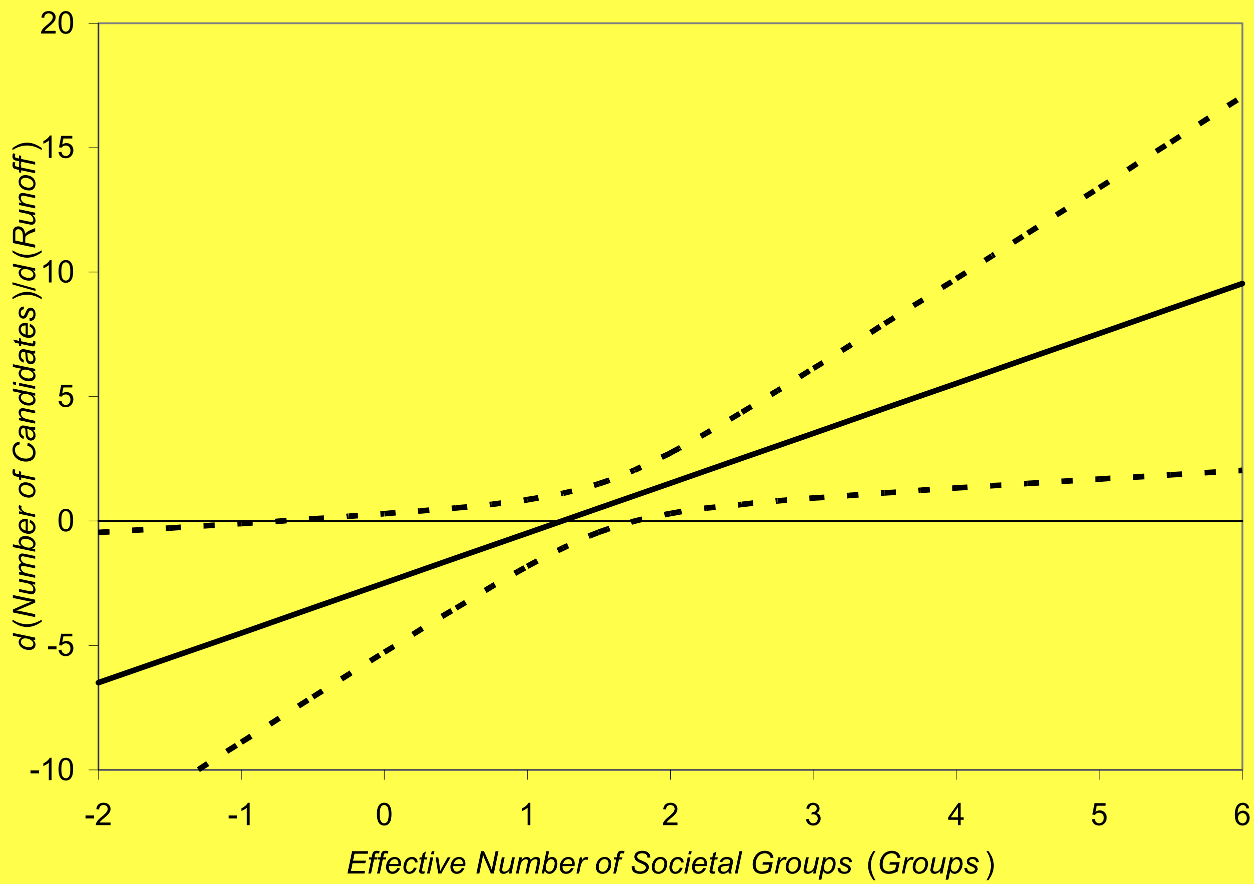
Use & Abuse of Some Common 'Rules'

- *Centering to Redress Colinearity Concerns*:
  - Adds no info, so changes ***nothing***; no help with colinearity or anything else; only moves substantive content of x=0,z=0.
  - Specifically, makes coeff. on x (z), effect when z (x) at sample-mean, the new 0. Do only if aids presentation.
- *Must Include All Components* (if $x{\cdot}z$, then x&z):
  - Application of Occam's Razor &/or scientific caution (e.g., greater flexibility to allow linear w/in lin-interax model), but
  - ***Not*** a logical or statistical requirement.
  - Safer rule than opposite & to check almost always, but
  - ***Not*** override *theory & evidence if (strongly) agree to exclude*
- *Pet-Peeve: Lingistic Gymnastics to Dodge the Math*
  - "Main effect, Interactive effect": ***the*** effect in model is dy/dx.
  - Discussion of [coefficients & s.e.'s] as if [effects & s.e.'s].

<u>Presentation</u>: Marginal-Effects / Differences Tables & Graphs

- Plot/Tabulate **Effects**, dy/dx, over Meaningful &/or Illuminating Ranges of z, with Conf. Int.'s
  - $d\hat{y}/dx \pm t_{df,p}\sqrt{Var(d\hat{y}/dx)} = \hat{\beta}_x + \hat{\beta}_{xz}z \pm t_{df,p}\sqrt{V(\hat{\beta}_x)+V(\hat{\beta}_{xz})z^2+2C(\hat{\beta}_x,\hat{\beta}_{xz})z}$
  - Explain axes
  - Explain shape
  - Linear-interax:
    - *Will* cross 0 & be insig @ 0.
  - Rescaling &
    - "main effect"
    - "centering"
    - Max(Asterisks)

Figure 5. Marginal Effect of *Runoff*, Extending the Range of *Groups*



Presentation: Expected-Value/Predictions Tables & Graphs

- ***Predictions***, $E(y|x,z)$:

$$\hat{y} \pm t_{df,p}\sqrt{Var(\hat{y})} =$$

$$\hat{\beta}_0 + \hat{\beta}_x x + \hat{\beta}_z z + \hat{\beta}_{xz} xz \pm t_{df,p}\sqrt{\begin{array}{l} V(\hat{\beta}_0) + V(\hat{\beta}_x)x^2 + V(\hat{\beta}_z)z^2 + V(\hat{\beta}_{xz})(xz)^2 \\ +2C(\hat{\beta}_0,\hat{\beta}_x)x + 2C(\hat{\beta}_0,\hat{\beta}_z)z + 2C(\hat{\beta}_0,\hat{\beta}_{xz})xz \\ +2C(\hat{\beta}_x,\hat{\beta}_z)xz + 2C(\hat{\beta}_x,\hat{\beta}_{xz})x^2 z + 2C(\hat{\beta}_z,\hat{\beta}_{xz})xz^2 \end{array}}$$

  - Here's one place a little matrix algebra would help:

$$\hat{y} \pm t_{df,p}\sqrt{Var(\hat{y})} = \hat{\beta}_0 + \hat{\beta}_x x + \hat{\beta}_z z + \hat{\beta}_{xz} xz \pm t_{df,p}\sqrt{\mathbf{x}'\hat{V}(\hat{\boldsymbol{\beta}})\mathbf{x}}$$

$$= \hat{\beta}_0 + \hat{\beta}_x x + \hat{\beta}_z z + \hat{\beta}_{xz} xz \pm t_{df,p}\sqrt{\begin{bmatrix} 1 & x & z & xz \end{bmatrix} \begin{bmatrix} \hat{V}(\hat{\beta}_0) & \hat{C}(\hat{\beta}_0,\hat{\beta}_x) & \hat{C}(\hat{\beta}_0,\hat{\beta}_z) & \hat{C}(\hat{\beta}_0,\hat{\beta}_{xz}) \\ \hat{C}(\hat{\beta}_0,\hat{\beta}_x) & \hat{V}(\hat{\beta}_x) & \hat{C}(\hat{\beta}_x,\hat{\beta}_z) & \hat{C}(\hat{\beta}_x,\hat{\beta}_{xz}) \\ \hat{C}(\hat{\beta}_0,\hat{\beta}_z) & \hat{C}(\hat{\beta}_x,\hat{\beta}_z) & \hat{V}(\hat{\beta}_z) & \hat{C}(\hat{\beta}_z,\hat{\beta}_{xz}) \\ \hat{C}(\hat{\beta}_0,\hat{\beta}_{xz}) & \hat{C}(\hat{\beta}_x,\hat{\beta}_{xz}) & \hat{C}(\hat{\beta}_z,\hat{\beta}_{xz}) & \hat{V}(\hat{\beta}_{xz}) \end{bmatrix} \begin{bmatrix} 1 \\ x \\ z \\ xz \end{bmatrix}}$$

  - Use spreadsheet or stat-graph software (…list coming…)

Presentation[1]: Choose Illuminating Graphics & Base Cases

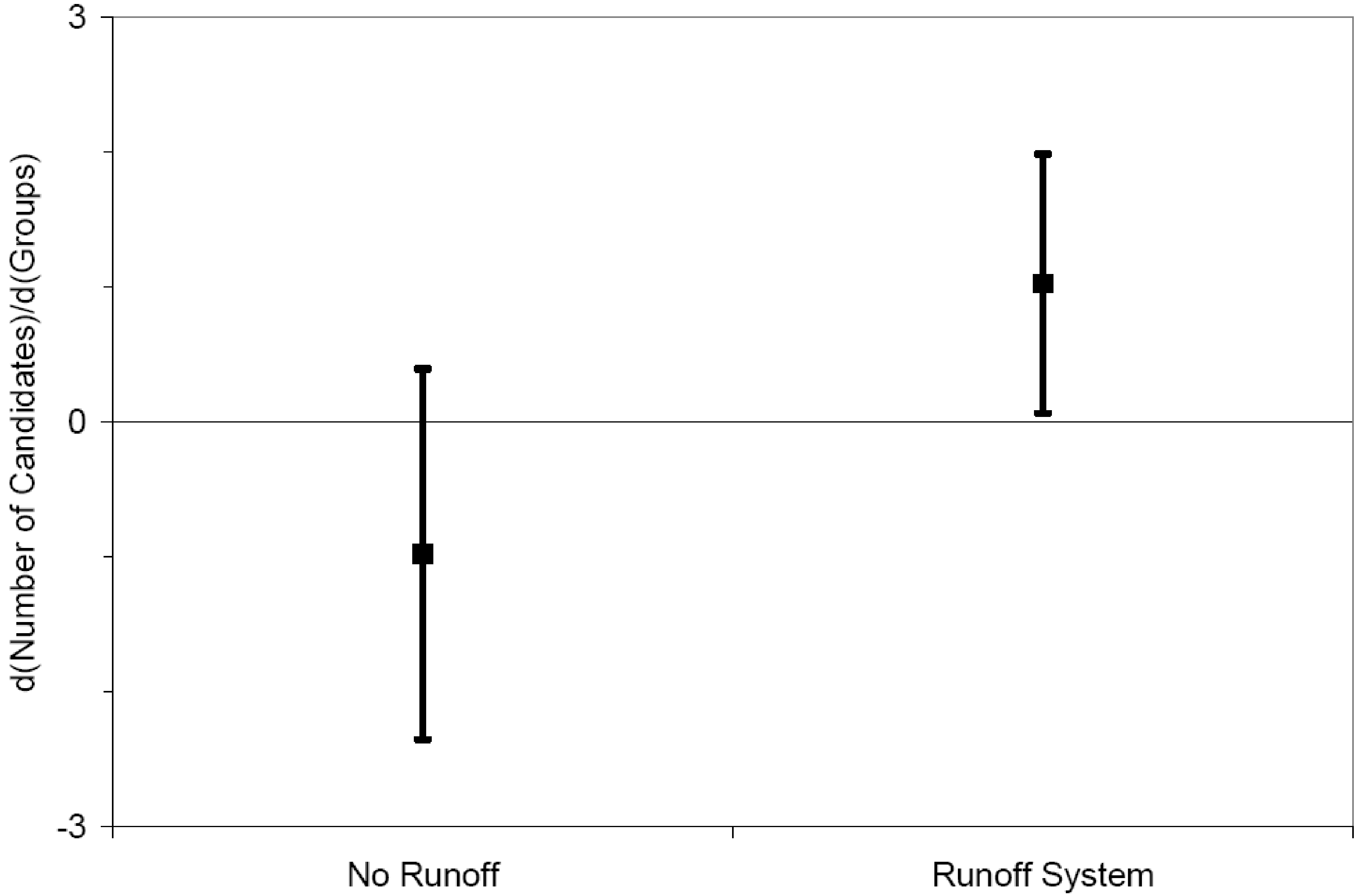
- Interpretation same regardless of "type" of interax: ***effect*** always ≡ $\boldsymbol{dy/dx}$, but present appropriately:
  - All combos Dummy, Discrete, or Continuous:
    - Dummy-Dummy=>4 (or $2^{\text{\#interacting variable}}$) points estimated, so box & whisker or histograms effective
    - Dummy-Continuous or Discrete(*few*)-Continuous=>2 (or # categories) slopes, so E(y|x,z) as line or dy/dx as box & whisker or histograms effective
    - Continuous – Continuous (or DiscMany)=>Effect-lines best or (slices from) contour plot (i.e., slices from 3D)
  - Powers (e.g., X & X^2=>parabola) viewable as interax w/ self; certain slope shifts too (e.g., $dy/dx=a$ for $x<x^0$ & b for $x>x^0$ is x interact w/ dummy for condition)

Presentation²:

Choose Illuminating Graphics & Base Cases

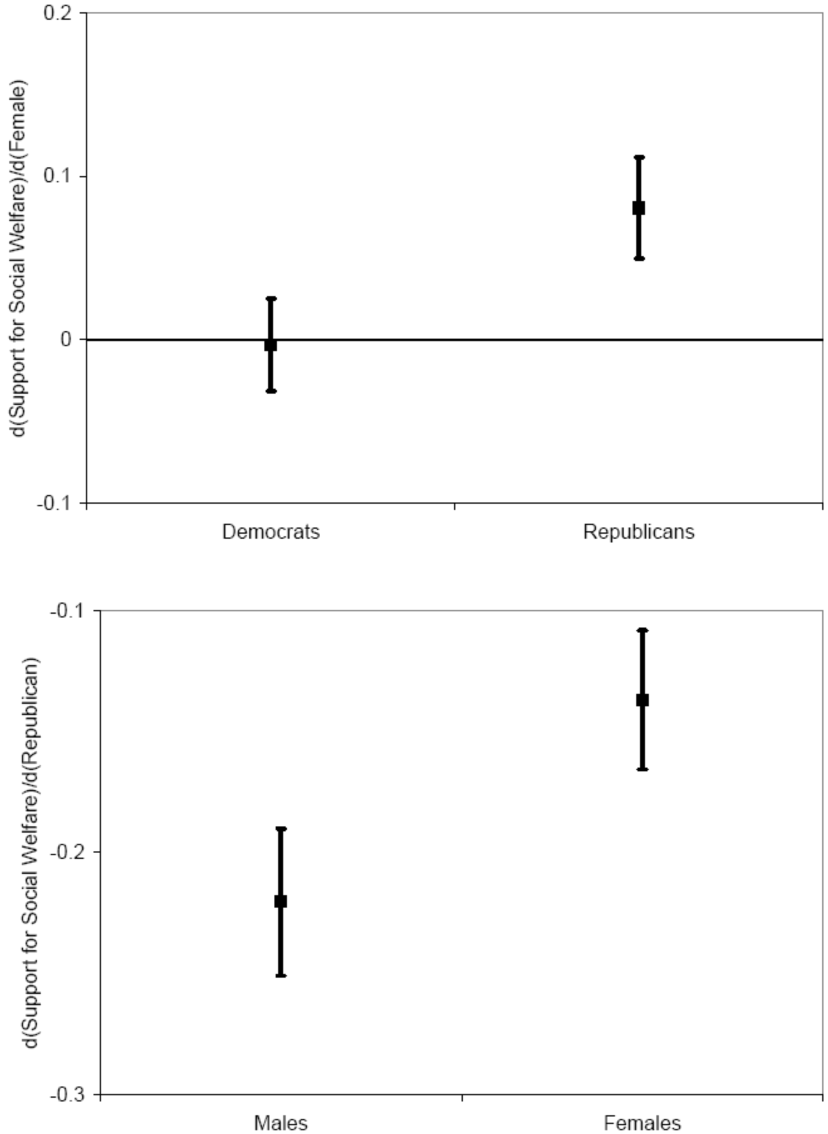
- Interpretation same regardless of "type" of interax: ***effect*** always ≡ $\boldsymbol{dy/dx}$, but present appropriately…
  - Always plot over substantively revealing ranges.
  - Especially with sets of dummies, have several (identical) specification options:
    - (full-set or set-less-1): choose which (& what base if use set-less-1) to abet presentation & discussion
    - (overlapping or disjoint): choose to facilitate presentation & discussion.
  - Scale Effectively: e.g., center only if & to extent that aids presentation & discussion (b/c centering does nothing else)

Presentation[3]: Choose Illuminating Graphics & Base Cases. Examples.



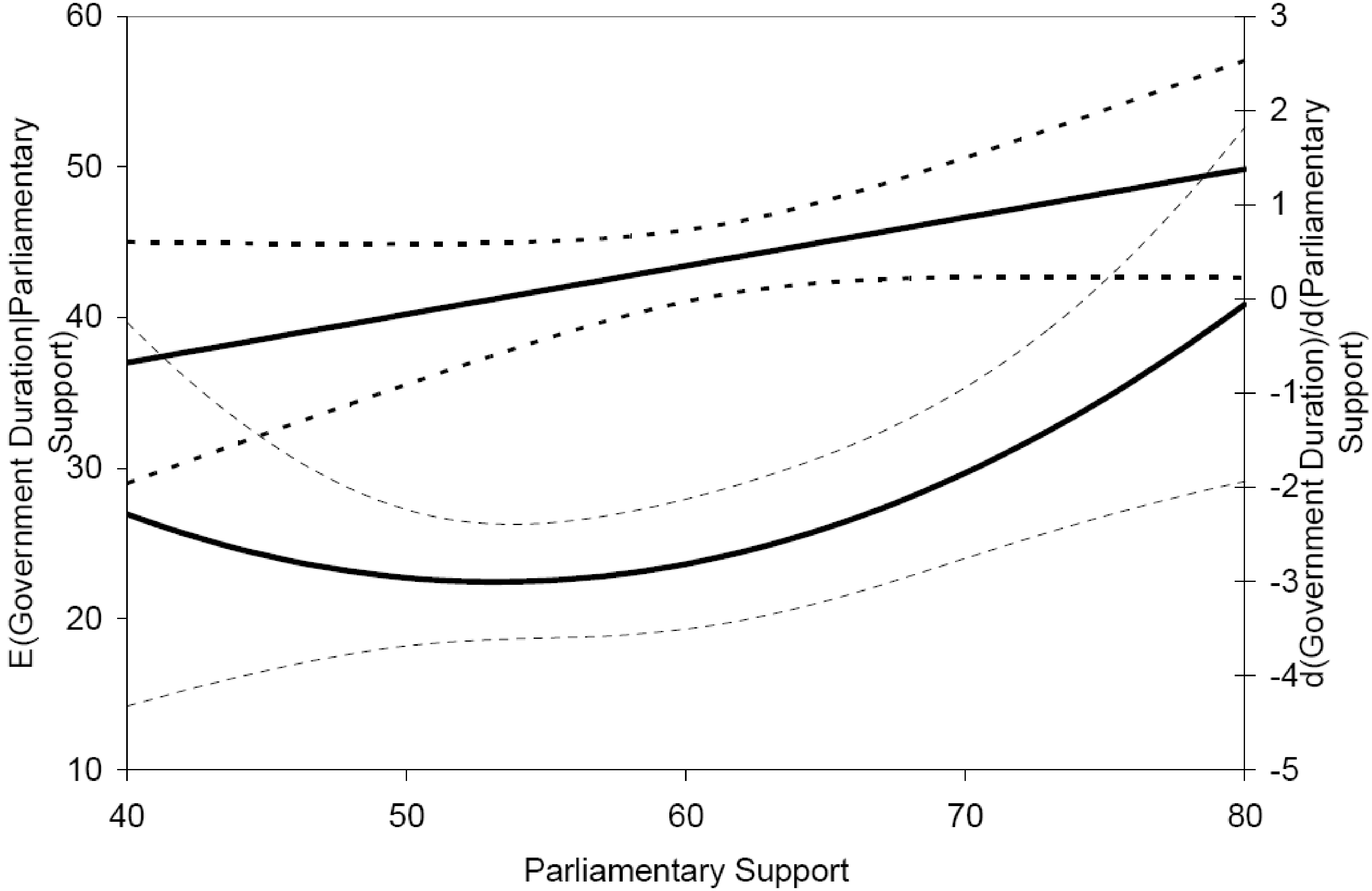
Dummy-Continuous Interaction: could also plot two E(Cands|Groups) lines, with c.i.'s, effectively.

Presentation[4]: Choose Illuminating Graphics & Base Cases. Examples.



Dummy-Dummy Interaction: could also plot four E(Supp|gender,party) box-whiskers effectively.

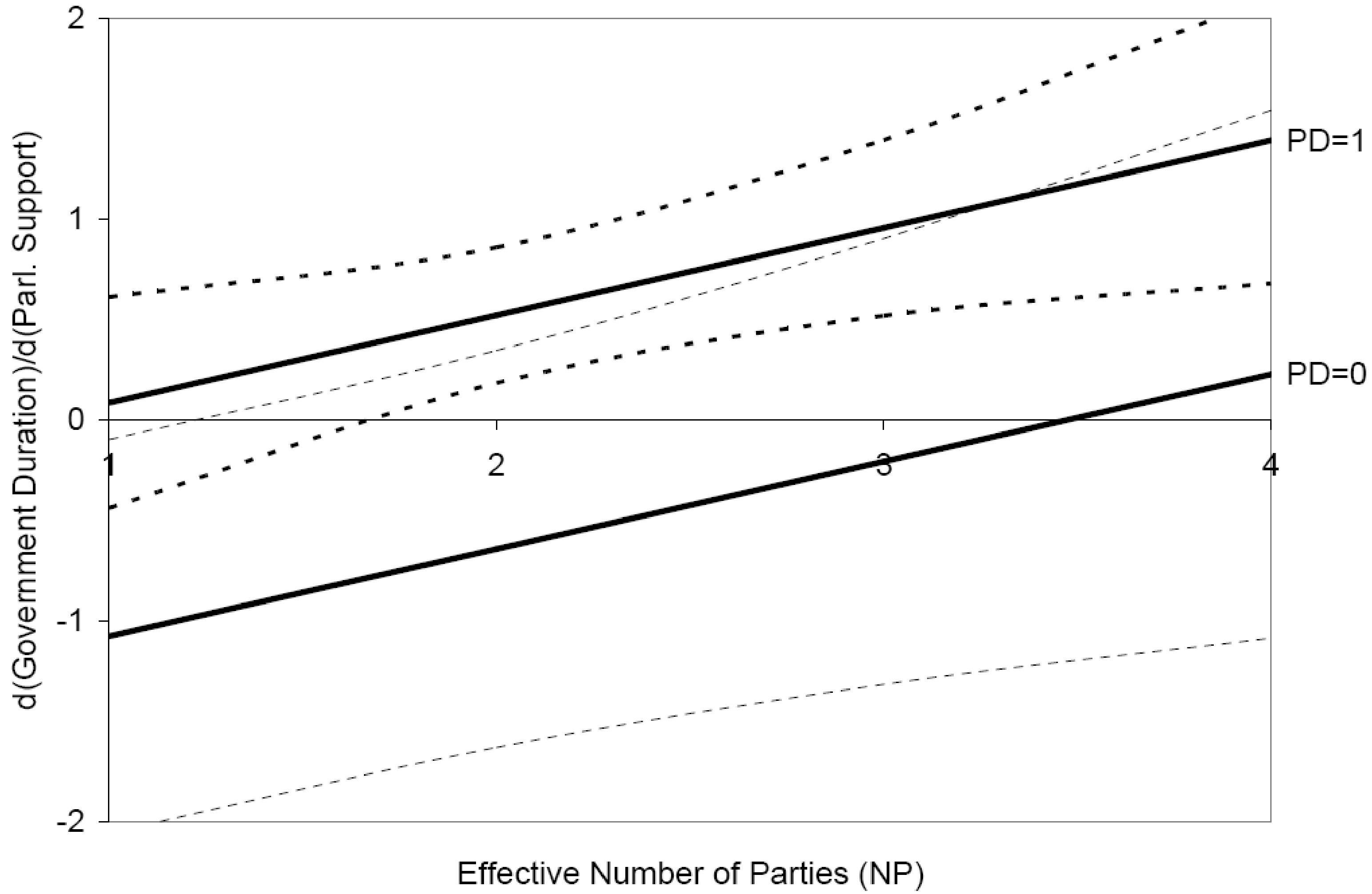
Presentation[5]: Choose Illuminating Graphics & Base Cases. Examples.



Quadratic Model of Government Duration as Function of Parliamentary Support of Governing Parties.

Presentation[6]: Choose Illuminating Graphics & Base Cases. Examples.

Figure 14. Marginal Effect of *Parliamentary Support for Government*, Pairwise-Interaction Model, with 90% Confidence Intervals



$$GovDur = \beta_0 + \beta_{np} NP + \beta_{ps} PS + \beta_{pd} PD + \beta_{npps} NP \times PS + \beta_{nppd} NP \times PD + \beta_{pdps} PD \times PS + \varepsilon \quad [25]$$

Elaborations, Complications, & Extensions: Sample-Splitting v. (Dummy-)Interacting[1]

- Split-sample (e.g., unit-by-unit) ≈ Full-Dum Interax:
  - Subsample by binary (or multinomial, e.g., CTRY in TSCS) category to estimate seperately ≈ Include dummy for each category (or set-less-1) & interact each dummy with each x *(*and include x by itself also if set-less-1*)*
    - Coeff's same (or equal substantive content if using *set-1* dummies).
    - S.E.'s same except s^2 part of OLS's $s^2(\mathbf{X'X})^{-1}$ is s_i^2 for splitting
    - Can make essentially exact by allow s_i^2 (FWLS)
  - Subsample by hi/lo values some non-nominal var equivalent to *nominalizing* the extra-nominal info & dummy-interact;
    - I.e., wasting information, when usually have too little (non-parametric or extreme-measurement-error arguments might justify)
    - So usually a bad idea…

Elaborations, Complications, & Extensions: Sample-Splitting v. (Dummy-)Interacting

- Split-sample abets eyeballing, obfuscates statistical analysis, of the main point: the different effects by category.
  - What's s.e/signif. of b_{1i}-b_{1j}? Need:

$$s.e.\left(b_{1i}-b_{1j}\right)=\sqrt{V\left(b_{1i}\right)+V\left(b_{1j}\right)-2C\left(b_{1i},b_{1j}\right)}=\sqrt{V\left(b_{1i}\right)+V\left(b_{1j}\right)}$$

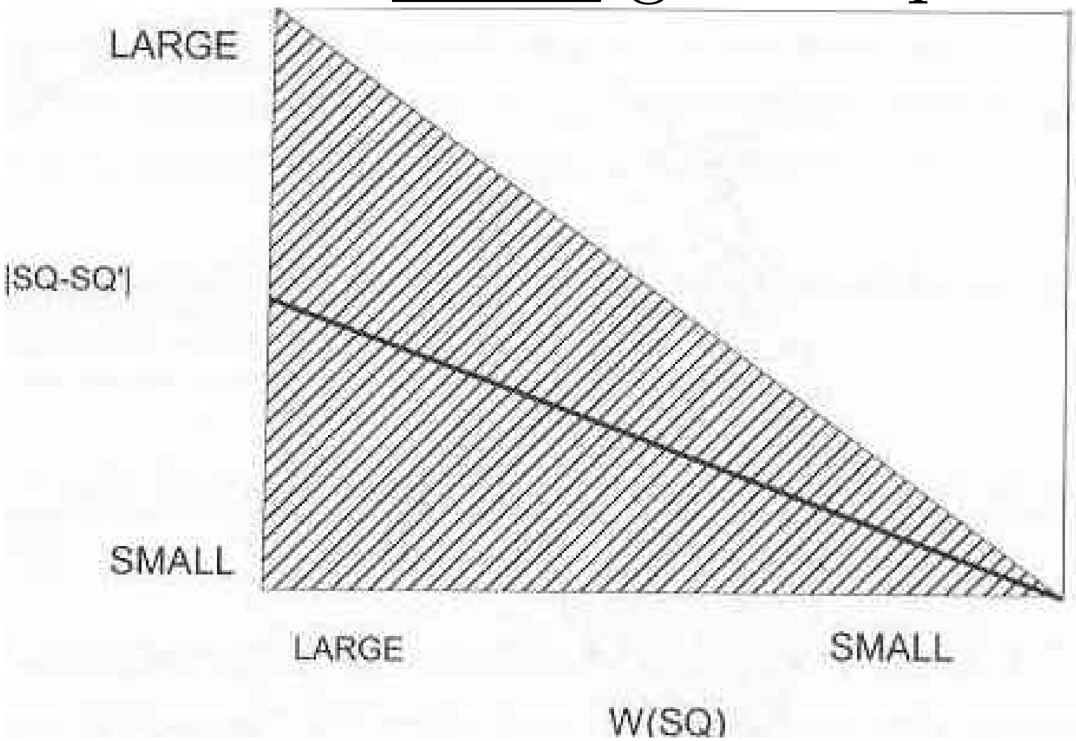
  - Luckily, cov=0, but, still, squaring 2 terms, sum, & root in head?
- Can choose *full dummy set* to mirror the split-sample estimates directly (& report that way, if wish) or the *set-less-one* to get significance of differences b/w samples directly (in the standard reported *t*-test)
  - Same thing, so choose for to optimize presentational efficacy.
- One advantage of hierarchical modeling is how it affords, naturally, various positions b/w these extremes.
  - E.g., can "borrow strength" across units.

Typical 2nd-Moment Implications of Interactions

- ***DMag*** permissive ele sys: ***allow*** more parties…

$$NP=\beta_0+\beta_1 DM+\varepsilon \;\;;\;\; V(\varepsilon)=f(DM) \;\;\text{, e.g., } \sigma_0+\sigma_1 DM$$

- Few ***Veto Actors*** ***allow*** greater policy-change…



$$y=\beta_0+\beta_1 VP+\varepsilon \;\;;\;\; V(\varepsilon)=f(VP) \;\;\text{, e.g., } \sigma_0+\sigma_1 VP$$

- I.e., these are Rndm-Coeff &/or Het-sked Props…

$$NP=\beta_0+\beta_1 DM+\beta_2 SF+\beta_3 DM\times SF+\varepsilon \;\;;\;\; V(\varepsilon)=f(DM) \;\;\text{, e.g., } \sigma_0+\sigma_1 DM$$

Sandwich Estimators

$$\boldsymbol{EN} = \beta_0 + \beta_1 \boldsymbol{SF} + \beta_2 \boldsymbol{DM} + \varepsilon$$

$$\partial \boldsymbol{EN} / \partial \boldsymbol{SF} = \beta_1 = \alpha_0 + \alpha_1 \boldsymbol{DM} + \omega_1$$

$$\partial \boldsymbol{EN} / \partial \boldsymbol{DM} = \beta_2 = \gamma_0 + \gamma_1 \boldsymbol{SF} + \omega_2$$

$$\Rightarrow \boldsymbol{EN} = \beta_0 + \left(\alpha_0 + \alpha_1 \boldsymbol{DM} + \omega_1\right) \boldsymbol{SF} + \left(\gamma_0 + \gamma_1 \boldsymbol{SF} + \omega_2\right) \boldsymbol{DM} + \varepsilon$$

$$= \beta_0 + \alpha_0 \boldsymbol{SF} + \left(\alpha_1 + \gamma_1\right) \boldsymbol{DM} \times \boldsymbol{SF} + \gamma_0 \boldsymbol{DM} + \left\{\varepsilon + \omega_1 \boldsymbol{SF} + \omega_2 \boldsymbol{DM}\right\}$$

$$= b_0 + b_1 \boldsymbol{SF} + b_2 \boldsymbol{DM} \times \boldsymbol{SF} + b_3 \boldsymbol{DM} + \varepsilon^*$$

- Notice the compound error term:
  - $V(\boldsymbol{\varepsilon}^*)$ will not be $\boldsymbol{\sigma}^2 \boldsymbol{I}$ even if $\boldsymbol{\varepsilon}$ is, so $V(\mathbf{b})$ doesn't reduce to $\boldsymbol{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, so OLS s.e.'s wrong.
  - Be OK on average (unbiased) & in limit (consistent) if element of $V(\boldsymbol{\varepsilon}^*)$ "orthogonal to $\mathbf{xx}'$"
  - But def'ly not because $\boldsymbol{\varepsilon}^*$ includes $\boldsymbol{x}$ & $\boldsymbol{z}$, which part of $\mathbf{X}$!

Sandwich Estimators

$$V(\mathbf{b}_{LS}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[V(\boldsymbol{\varepsilon} + \omega_1\mathbf{S} + \omega_2\mathbf{D}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- Brilliant insight of ‘robust’ (i.e., consistent) “sandwich” estimators:
  - Only need formula that accounts relation V($\boldsymbol{\varepsilon}^*$) to “$\mathbf{X}'\mathbf{X}$”, i.e., regressors, squares, & cross-prod’s involved in $\mathbf{X}'[\cdot]\mathbf{X}$”
- $\Rightarrow$ “, robust” (or, in HM: “, cluster”) can work:

$V(\varepsilon_i^*)_{RE} = \sigma_\varepsilon^2 + \sigma_{\omega_1}^2 x_i^2 + \sigma_{\omega_2}^2 z_i^2$ so track e^2 rel $\boldsymbol{xx'}$ & $\mathbf{zz}'$ $\Rightarrow$ White’s *heteroskedasticity-consistent s.e.’s*:

$$[\cdot] = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

Cross-Level Interactions

- At this stage, nothing much different (n.b., not many subscripts used above):
  - If CLRM assumptions apply, then unbiased, consistent, and efficient.
    - Two main issues of concern:
      - Parameter heterogeneity (incl. intercept): this can cause bias if pattern unmodeled heterogeneity relates to **X**,
      - Non-spherical error cov-mat: an efficiency & proper s.e.'s issue, not a bias/consistency one
  - As before: Effects, their variances, symmetry of interactive prop's, that neither *micro-* nor *macro-level* coefficients=effects.

Cross-Level Interactions:

```
reg spend L.spend unem left growthpc depratio
cdem trade lowwage fdi skand skand_unem
```

$$\text{spend}_{it} = \beta_i^0 + \beta_i^l \text{left}_{it} + \ldots + \varepsilon_{it}$$

$$\beta_i^0 = \alpha_0 + \alpha_1 \text{skand}_i + u_i^0$$

$$\beta_i^l = \gamma_0 + \gamma_1 \text{skand}_i + u_i^1$$

$$\Rightarrow \text{spend}_{it} = \alpha_0 + \alpha_1 \text{skand}_i + u_i^0 + \gamma_0 \text{left}_{it} + \gamma_1 \text{left}_{it} \times \text{skand}_i + \text{left}_{it} u_i^1 + \ldots + \varepsilon_{it}$$

gathering terms :

$$\text{spend}_{it} = \alpha_0 + \ldots + \alpha_1 \text{skand}_i + \gamma_0 \text{left}_{it} + \gamma_1 \text{left}_{it} \times \text{skand}_i + \left(u_i^0 + \text{left}_{it} u_i^1 + \varepsilon_{it} \right)$$

$$\Rightarrow \frac{\partial \text{spend}}{\partial \text{left}} = b_{left} + b_{lftsk} \text{skand} + \text{u}_i^1 \quad \& \quad \frac{\partial \text{spend}}{\partial \text{skand}} = b_{skand} + b_{lftsk} \text{left}$$

$$V(\varepsilon_i^*)_{HM} = \sigma_0^2 + \sigma_1^2 \boldsymbol{x}_i \boldsymbol{x}_i' + \sigma_2^2 \boldsymbol{z}_i \boldsymbol{z}_i'$$

$\Rightarrow$ *cluster:*

$$[\cdot] = \sum_{j=1}^{J} \left\{ \left(\sum_{i=1}^{n_j} e_{ij} \boldsymbol{x}_{ij}\right)' \left(\sum_{i=1}^{n_j} e_{ij} \boldsymbol{x}_{ij}\right) \right\}$$

From CLRM to Multilevel Model

$$\text{spend}_{it} = \beta_i^0 + \beta_i^l \text{left}_{it} + \ldots + \varepsilon_{it}$$

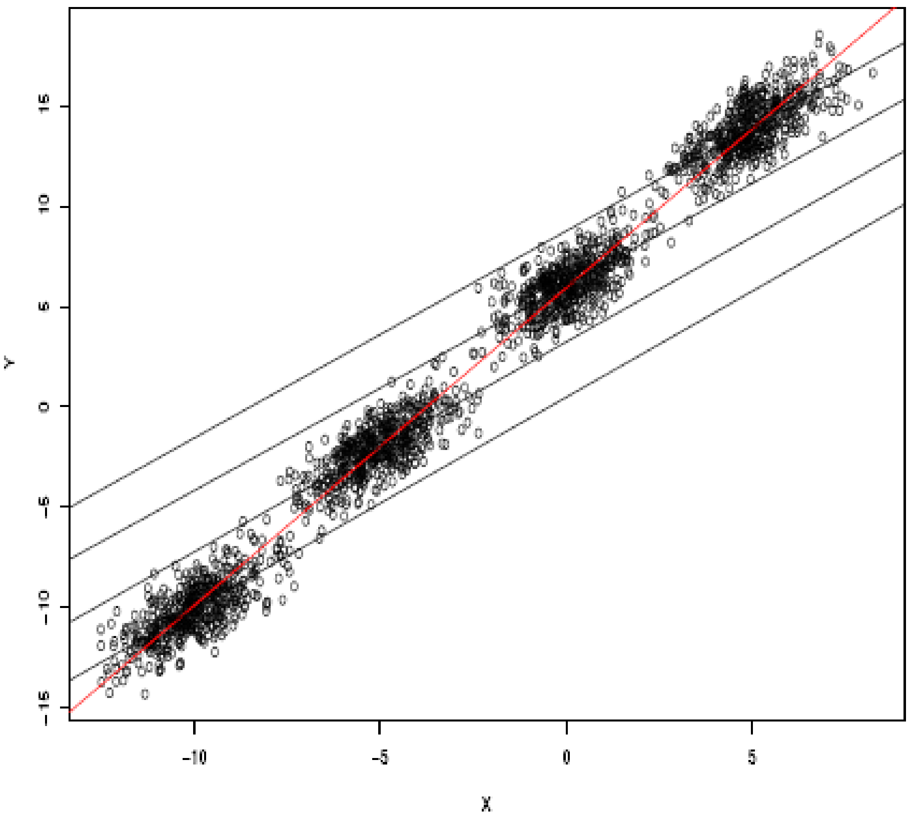
$$\beta_i^0 = \alpha_0 + \alpha_1 \text{skand}_i + u_i^0$$

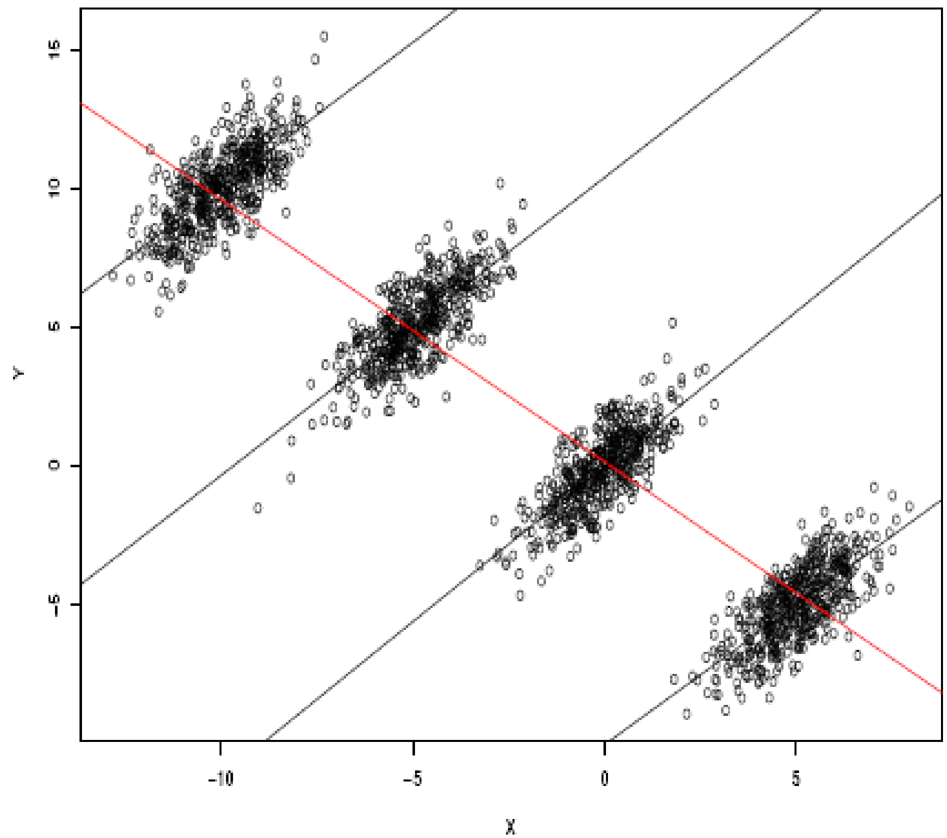
$$\beta_i^l = \gamma_0 + \gamma_1 \text{skand}_i + u_i^1$$

- If CLRM assumptions apply, then OLS unbiased, consistent, and efficient.
  - Two main issues of concern:
    - Parameter heterogeneity: (see pictures)
      - systematic &/or stochastic (fixed v. rndm intrcpt/coeff)
      - can cause bias if pattern unmodeled hetero relates to **X**,
    - Non-spherical error cov-mat: an efficiency & proper s.e.'s issue, not a bias/consistency one
      - But "mere inefficiency" can be serious.
      - And accurate std err's very important.

From the CLRM to HLM

- Examples of parameter hetereogeneity that covaries w/ X values, so bias:





  - Note: FE v. RE both theoretically could cause bias if cov w/ X, but latter i.d.'d off orthog

From the CLRM to RE Model

- Std. R.E. Model: Odd that std. lin-interax model:
  - Assumes know *y=f(X)+error*: $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_{xz} x_i z_i + \varepsilon_i$
  - But *dy/dx=f(z)* w/o error!: $\frac{\partial y}{\partial x} = \beta_x + \beta_{xz} xz$
  - So, try:

$$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon^0$$
$$\frac{\partial y}{\partial x} \equiv \beta_1 = \alpha_0 + \alpha_1 z + \varepsilon^1$$
$$\frac{\partial y}{\partial z} \equiv \beta_2 = \gamma_0 + \gamma_1 x + \varepsilon^2$$
$$\Rightarrow y = \beta_0 + \left(\alpha_0 + \alpha_1 z + \varepsilon^1\right) x + \left(\gamma_0 + \gamma_1 x + \varepsilon^2\right) z + \varepsilon^0$$
$$= \beta_0 + \alpha_0 x + \gamma_0 z + \left(\alpha_1 + \gamma_1\right) xz + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)$$

  - => std. lin-interact...except compound error-term…
- Std. HLM: Same model, except x_{ij} & z_j, &
  - So std lin-interact, but w/ diff compound-error struct.

$$\varepsilon^* = \varepsilon_{ij}^0 + \varepsilon_j^1 x_{ij} + \varepsilon_{ij}^2 z_j$$

From CLRM to Hierarchical Model

- Std. HLM: Same model, except x_{ij} & z_j, &
  - So a std. lin-interact too, but with different compound-error stochastic properties.

$$\text{spend}_{it} = \beta_i^0 + \beta_i^l \text{left}_{it} + \ldots + \varepsilon_{it}$$

$$\beta_i^0 = \alpha_0 + \alpha_1 \text{skand}_i + u_i^0$$

$$\beta_i^l = \gamma_0 + \gamma_1 \text{skand}_i + u_i^1$$

$$\Rightarrow \text{spend}_{it} = \alpha_0 + \alpha_1 \text{skand}_i + u_i^0 + \gamma_0 \text{left}_{it} + \gamma_1 \text{left}_{it} \times \text{skand}_i + \text{left}_{it} u_i^1 + \ldots + \varepsilon_{it}$$

gathering terms:

$$\text{spend}_{it} = \alpha_0 + \ldots + \alpha_1 \text{skand}_i + \gamma_0 \text{left}_{it} + \gamma_1 \text{left}_{it} \times \text{skand}_i + \left(u_i^0 + \text{left}_{it} u_i^1 + \varepsilon_{it} \right)$$

$$\Rightarrow \frac{\partial \text{spend}}{\partial \text{left}} = b_{left} + b_{lftsk} \text{skand} + \text{u}_i^1 \quad \& \quad \frac{\partial \text{spend}}{\partial \text{skand}} = b_{skand} + b_{lftsk} \text{left}$$

Properties of OLS under HLM Conditions

- Properties of OLS Estimates of Lin-Interact Model if truly RE/HLM:

$$y = \beta_0 + \beta_x x + \beta_z z + \beta_{xz} xz + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \mathbf{X\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)$$

- So, OLS coeff. est's sill differ from truth by $\mathbf{A}\boldsymbol{\varepsilon}^*$:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathbf{LS}} &= (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}\left[\mathbf{X\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= (\mathbf{X'X})^{-1}\mathbf{X'X\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon^*\end{aligned}$$

- So, OLS coeff. est's unbiased & consistent:

$$\begin{aligned}E\left(\hat{\boldsymbol{\beta}}_{\mathbf{LS}}\right) &= E\left[\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon^*\right] = E\left[\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}E\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left[E\left(\varepsilon^0\right) + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] \\ &= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left[0 + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left[0+0+0\right] = \boldsymbol{\beta}. \quad Q.E.D.\end{aligned}$$

  – Note: only works for models w/ additively separable stochastic component; not nec'ly for others (e.g., logit/probit)

Properties of OLS under HLM Conditions

- But, OLS s.e.'s will be wrong; not $s^2(\mathbf{X'X})^{-1}$, but:

$$
\begin{aligned}
V\left(\hat{\boldsymbol{\beta}}_{\mathbf{LS}}\right) &= V\left[\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}^*\right] \\
&= V[\boldsymbol{\beta}] + V\left[(\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}^*\right] + 2C\left[\boldsymbol{\beta}, (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}^*\right] \\
&= 0 + V\left[(\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}^*\right] + 0 \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}V\left(\boldsymbol{\varepsilon}^*\right)\mathbf{X}(\mathbf{X'X})^{-1} \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}\left[V\left(\boldsymbol{\varepsilon}^0 + \boldsymbol{\varepsilon}^1\boldsymbol{x} + \boldsymbol{\varepsilon}^2\boldsymbol{z}\right)\right]\mathbf{X}(\mathbf{X'X})^{-1} \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}\left[V\left(\boldsymbol{\varepsilon}^0\right) + V\left(\boldsymbol{\varepsilon}^1\boldsymbol{x}\right) + V\left(\boldsymbol{\varepsilon}^2\boldsymbol{z}\right)\right]\mathbf{X}(\mathbf{X'X})^{-1}
\end{aligned}
$$

(the covariance terms are assumed zero)

Sandwich Estimators

$$V\left(\hat{\boldsymbol{\beta}}_{\mathbf{LS}}\right) = (\mathbf{X'X})^{-1}\mathbf{X'}\left[V\left(\boldsymbol{\varepsilon}^{0}\right) + V\left(\boldsymbol{\varepsilon}^{1}\boldsymbol{x}\right) + V\left(\boldsymbol{\varepsilon}^{2}\boldsymbol{z}\right)\right]\mathbf{X}(\mathbf{X'X})^{-1}$$

- Not $\sigma^2 \boldsymbol{I}$ (even if each $\boldsymbol{\varepsilon}^*$ component is), so whole thing doesn't reduce to $\sigma^2(\mathbf{X'X})^{-1}$, so OLS s.e.'s wrong.
- Be OK on avg (unbiased) & in limit (consistent) if that term varied in way " orthogonal to **XX'** "
  - But def'ly not b/c [·] includes ***x*** & ***z***, which part of **X**!
  - =brilliant insight of 'robust' (i.e., consistent) s.e. est's:
    - Only need s.e. formula that accounts relation V($\boldsymbol{\varepsilon}^*$) to "**X'X**", i.e., regressors, squares, & cross-prod's involved in **X'[·]X**"
- ⇒ ", robust" & ", cluster" can work (for RE & HLM, resp'ly)
  - $\hat{V}\left(\hat{\boldsymbol{\beta}}\right)_{RE} = \sigma^2\left(\mathbf{I} + \boldsymbol{xx'} + \boldsymbol{zz'}\right)$ so track e^2 rel ***xx'*** & **zz'** ⇒ $[\cdot] = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{X}_i \boldsymbol{X}_i'$
    - i.e., White's het-consistent s.e.'s
  - $\hat{V}\left(\hat{\boldsymbol{\beta}}\right)_{HM} = \sigma_0^2\mathbf{I} + \sigma_1^2\boldsymbol{xx'} + \sigma_2^2\boldsymbol{zz'}$ sim but grpng ⇒ $[\cdot] = \sum_{j=1}^{J}\left\{\left(\sum_{i=1}^{n_j} e_{ij}\boldsymbol{X}_{ij}\right)'\left(\sum_{i=1}^{n_j} e_{ij}\boldsymbol{X}_{ij}\right)\right\}$
    - i.e., het-cluster consistent s.e.'s

From the CLRM to HLM

$$y = \beta_0 + \alpha_0 x + \gamma_0 z + (\alpha_1 + \gamma_1) xz + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)$$

$$V(\hat{\boldsymbol{\beta}}_{LS}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[V(\varepsilon^0) + V(\varepsilon^1 x) + V(\varepsilon^2 z)]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- $\Rightarrow$ appropriate ", robust" & ", cluster" can work
  - I.e., <u>*asymptotically*</u> std errs right…BUT need large n_j
  - I.e., <u>*coefficients still inefficient*</u>.
    - Want/need efficiency, or n_j low? HLM/RE or FGLS/FWLS.
    - Note: similarity RE and HLM, RE & FWLS. As suggests, RE only helps efficiency and only rightly does so if that's all it does. (I.e., if the RE's orthogonal to X.)
  - I.e., "work" thusly for models with additively-separable stochastic components
    - As w/ all such "sandwich" estimators, logical disconnect in applying them to models w/o such separability.