Empirical Models of Context Conditionality, (Inter)Dependence, and Endogeneity

Robert J. Franzese, Jr.
Professor of Political Science
The University of Michigan, Ann Arbor
franzese@umich.edu
http://www.umich.edu/~franzese

Essex Summer School in Social Science Data Collection & Analysis

10-21 August 2009

Introductions and an Introduction

I. Introductions…

II. Syllabus & Lecture+Lab Materials… (+/- daily posting to web)

1a. Introduction: The Centrality of Model Specification to Empirical Analysis

1b. Reviews:
- Matrix Algebra & Basic Calculus
- Classical (Normal) Linear-Regression Model

2a. Reviews:
- Generalized (Normal) Linear-Regression Model
- Maximum Likelihood & Nonlinear/Qualitative/Limited Dependent-Variable Models

2b. The Centrality of Model Specification to Empirical Analysis: Reprise

3. Linear-Interaction Models

4. Nonlinear Interaction Models

5. Multilevel Interaction Models

6. Temporally Dynamic Models

7. Spatially and Spatiotemporally Dynamic Models I

8. Spatially and Spatiotemporally Dynamic Models II

9. Systems of Endogenous Equations

10. Nonlinear Endogenous Systems

III. *Introduction*:

Context Matters—The EMTI™ *or Model It*™ *Strategy*

A. Fundamental Challenges to Empirical Evaluation in the Social Sciences

1. **Multicausality**: Just about everything matters…
2. **Context Conditionality**: how just about everything matters depends on just about everything else…
3. **Endogeneity**: just about everything is endogenous to just about everything else…
4. (**Micronumerosity**: …and we usually have far too little empirical information to figure it all out.)

B. Options for ways forward:

1. **Qualitative Analysis**: Not going to help *for empirical evaluation* because…
   a) …if many things matter, generally too much moving in any small number of contexts,
   b) …and world likely partly stochastic, or at least relative to systematic part in our theories,
   c) …& *controlling* moving parts qualitatively beyond human capability, even 2 parts with constant, not context-conditional, linear-additive & separable effects is beyond me:

$$Min_{b_1,b_2}\sum_{i=1}^{n}\left(Dem_i-b_1EcDev_i-b_2EqDist_i\right)^2 \Rightarrow \begin{cases} \text{(i)} \quad \dfrac{\partial\sum_{i=1}^{n}\left(Dem_i-b_1EcDev_i-b_2EqDist_i\right)^2}{\partial b_1}=0 \\ \text{(ii)} \quad \dfrac{\partial\sum_{i=1}^{n}\left(Dem_i-b_1EcDev_i-b_2EqDist_i\right)^2}{\partial b_2}=0 \end{cases}$$

$$\text{(i)} \;\Rightarrow \begin{cases} \sum_{i=1}^{n}EcDev_i\left(Dem_i-b_1EcDev_i-b_2EqDist_i\right)=0 \Rightarrow \sum_{i=1}^{n}EcDev_iDem_i=b_1\sum_{i=1}^{n}EcDev_i^2+b_2\sum_{i=1}^{n}EcDev_iEqDist_i \\ \Rightarrow b_1=\left(\sum_{i=1}^{n}EcDev_iDem_i-b_2\sum_{i=1}^{n}EcDev_iEqDist_i\right)\Big/\sum_{i=1}^{n}EcDev_i^2 \\ \Rightarrow b_1=\dfrac{Cov\left(Dem,EcDev\right)-b_2\times Cov\left(EcDev,EqDist\right)}{Var(EcDev)} \end{cases}$$

$$\text{(ii)} \;\Rightarrow\; \text{(analogously)} \begin{cases} b_2=\left(\sum_{i=1}^{n}EqDist_iDem_i-b_1\sum_{i=1}^{n}EcDev_iEqDist_i\right)\Big/\sum_{i=1}^{n}EqDist_i^2 \\ b_2=\dfrac{Cov\left(Dem,EqDist\right)-b_1\times Cov\left(EcDev,EqDist\right)}{Var(EqDist)} \end{cases}$$

$$\Rightarrow \begin{cases} b_1=\dfrac{\left(\sum_{i=1}^{n}EcDev_iDem_i\right)\left(\sum_{i=1}^{n}EqDist_i^2\right)-\left(\sum_{i=1}^{n}EqDist_iDem_i\right)\left(\sum_{i=1}^{n}EcDev_iEqDist_i\right)}{\left(\sum_{i=1}^{n}EcDev_i^2\right)\left(\sum_{i=1}^{n}EqDist_i^2\right)-\left(\sum_{i=1}^{n}EcDev_iEqDist_i\right)^2} \\ b_1=\dfrac{Cov\left(Dem,EcDev\right)\times Var(EqDist)-Cov\left(Dem,EqDist\right)\times Cov\left(EcDev,EqDist\right)}{Var(EcDev)\times Var(EqDist)-\left[Cov\left(EcDev,EqDist\right)\right]^2} \end{cases}$$

and, analogously:

$$\Rightarrow \begin{cases} b_2=\dfrac{\left(\sum_{i=1}^{n}EqDist_iDem_i\right)\left(\sum_{i=1}^{n}EcDev_i^2\right)-\left(\sum_{i=1}^{n}EcDev_iDem_i\right)\left(\sum_{i=1}^{n}EcDev_iEqDist_i\right)}{\left(\sum_{i=1}^{n}EcDev_i^2\right)\left(\sum_{i=1}^{n}EqDist_i^2\right)-\left(\sum_{i=1}^{n}EcDev_iEqDist_i\right)^2} \\ b_2=\dfrac{Cov\left(Dem,EqDist\right)\times Var(EcDev)-Cov\left(Dem,EcDev\right)\times Cov\left(EcDev,EqDist\right)}{Var(EcDev)\times Var(EqDist)-\left[Cov\left(EcDev,EqDist\right)\right]^2} \end{cases}$$

d) …and, of course, if the effect varies depending on context, then obviously…

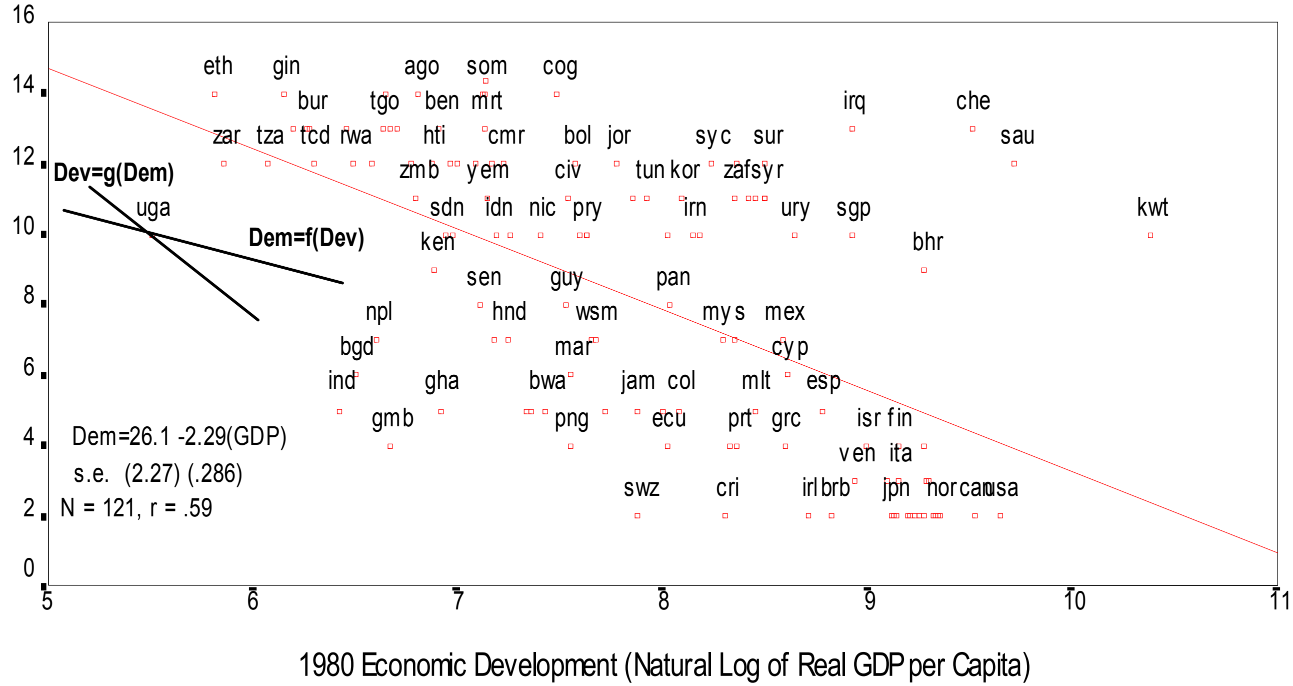
(1) …you need more contexts, and

(2) …you need more mental capacity for juggling variances & covariances of crossproducts too!

e) …and nothing about looking more closely *per se* addresses the endogeneity problem…

(1) …of course, for that matter, nothing about looking more summarily & systematically at many contexts *per se* addresses the endogeneity problem either. The problem is:

The Relationship Between the
Degree of Democracy and of Economic Development



$$Y = a \times X; \quad X = b \times Y$$

$$\Rightarrow Y = a \times (b \times Y)$$

$$\Rightarrow ab = 1 \text{ or } a = b^{-1}$$

(2) …you don't need more empirical detail or more empirical contexts; you need more structure! For instance, consider a system of M endogenous equations like this:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}_i' \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{bmatrix}_i' \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{km} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_m \end{bmatrix}_i$$

In matrix notation, the system may be written compacts as: $\mathbf{y}_i'\mathbf{\Gamma} + \mathbf{x}_i'\mathbf{B} = \mathbf{\varepsilon}_i$. Even just $V(\varepsilon) \equiv \mathbf{\Sigma}$ has $½M^2+½M > M$ things to learn, in general, from each M things observed in context i….

f) …and the terms of the sacrifice *quantity-for-quality* tradeoff not usually appealing…

(1) Model with measurement error in DepVar:

$$QualDem = QualDem^* + \gamma \Rightarrow QualDem = \beta \times EcDev + \varepsilon - \gamma$$

(2) Under the usual conditions, then:

$$\hat{\beta} = \frac{Cov(QualDem, EcDev)}{Var(EcDev)} \quad \text{and} \quad V\left(\hat{\beta}\right) = \frac{\sigma_\varepsilon^2 + \sigma_\gamma^2}{Var(EcDev)} = \frac{\sigma_\varepsilon^2 + \sigma_\gamma^2}{(n-1) \times \sigma_x^2}$$

(3) In the following table, the *quality-quantity tradeoff* is read by comparing going up rows (quantity reduction) to going right-to-left across columns (quality improvement). Note that quantity reduction will also often imply movements up the blocks within the table.

Variation of Explanatory Variable (σ_x^2): 0.5							
	σ_γ^2=0.1	σ_γ^2=0.25	σ_γ^2=0.5	σ_γ^2=1	σ_γ^2=2	σ_γ^2=4	σ_γ^2=10
n=2	2.200	2.500	3.000	4.000	6.000	10.000	22.000
N=10	0.244	0.278	0.333	0.444	0.667	1.111	2.444
N=50	0.045	0.051	0.061	0.082	0.122	0.204	0.449
N=250	0.009	0.010	0.012	0.016	0.024	0.040	0.088
Variation of Explanatory Variable (σ_x^2): 1							
	σ_γ^2=0.1	σ_γ^2=0.25	σ_γ^2=0.5	σ_γ^2=1	σ_γ^2=2	σ_γ^2=4	σ_γ^2=10
n=2	1.100	1.250	1.500	2.000	3.000	5.000	11.000
N=10	0.122	0.139	0.167	0.222	0.333	0.556	1.222
N=50	0.022	0.026	0.031	0.041	0.061	0.102	0.224
N=250	0.004	0.005	0.006	0.008	0.012	0.020	0.044
Variation of Explanatory Variable (σ_x^2): 2							
	σ_γ^2=0.1	σ_γ^2=0.25	σ_γ^2=0.5	σ_γ^2=1	σ_γ^2=2	σ_γ^2=4	σ_γ^2=10
n=2	0.550	0.625	0.750	1.000	1.500	2.500	5.500
N=10	0.061	0.069	0.083	0.111	0.167	0.278	0.611
N=50	0.011	0.013	0.015	0.020	0.031	0.051	0.112
N=250	0.002	0.003	0.003	0.004	0.006	0.010	0.022

(4) …and that's *per parameter* (relationship) to be estimated!!

2. **<u>"Robust" Estimation Strategies</u>**: Not ideal *for soc sci emp eval* because…

a) …strategies that relax structural impositions generally do so at cost of efficiency: need more info, and fundamental problem 0 tends to be too little info…

b) …nonparametric (i.e., less structural) mechanisms for allowing context-conditionality & ubiquitous endogeneity tend, paradoxically, to be proliferation of parameters to estimate.

c) …assumptions don't fit: SUTVA, separability, deterministic causation…

d) …in fact, purely nonparametric not possible; very easy to show:

e) One way to see that:

$$y_{it} = f_{it}\left(\mathbf{x}_{it}, \boldsymbol{\beta}_{it}, \varepsilon_{it}\right); \;\; \boldsymbol{\varepsilon} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{it}); \;\; i = 1..N, \; t = 1..T, n = NT$$

has K β's plus $\frac{1}{2}(NT)^2 + \frac{1}{2}NT$ (unique parameters in each v-cov matrix for ε) total parameters ***<u>per function</u>*** to learn ***<u>from each information-set observed</u>***.

f) Another way to see it: even just v-cov mat, even if assumed constant, has $\frac{1}{2}(NT)^2 + \frac{1}{2}NT > NT$ potentially unique parameters in it:

$$\mathbf{V}\left(\boldsymbol{\varepsilon}\right) \equiv \mathbf{V}\left(\mathbf{y} \mid \mathbf{X}\right) \equiv \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega} =$$

$$\sigma^2 \times \begin{bmatrix}
\omega_{1,1}^2 & \omega_{1,12} & \omega_{1,13} & \cdots & \omega_{1,1T} & \omega_{12,11} & \omega_{12,12} & \omega_{12,13} & \cdots & \omega_{12,1T} \\
\omega_{1,21} & \omega_{1,2}^2 & & & \vdots & \omega_{12,21} & \omega_{12,22} & & & \vdots \\
\omega_{1,31} & & \omega_{1,3}^2 & & \vdots & \omega_{12,31} & & \omega_{12,33} & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\
\omega_{1,T1} & \cdots & \cdots & \cdots & \omega_{1,T}^2 & \omega_{12,T1} & \cdots & \cdots & \cdots & \omega_{12,TT} \\
\omega_{21,11} & \omega_{21,12} & \omega_{21,13} & \cdots & \omega_{21,1T} & \omega_{2,1}^2 & \omega_{2,12} & \omega_{2,13} & \cdots & \omega_{2,1T} \\
\omega_{21,21} & \omega_{21,22} & & & \vdots & \omega_{2,21} & \omega_{2,2}^2 & & & \vdots \\
\omega_{21,31} & & \omega_{21,33} & & \vdots & \omega_{2,31} & & \omega_{2,3}^2 & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\
\omega_{21,T1} & \cdots & \cdots & \cdots & \omega_{21,TT} & \omega_{2,T1} & \cdots & \cdots & \cdots & \omega_{2,T}^2
\end{bmatrix}$$

3. The **EMTI**™ or ***Model It***™ strategy: Lean harder on theory/substance to specify $E(\mathbf{y}|\mathbf{X})$ more precisely—nature of interactions, functional forms, measures, etc.

   a) Refines question put to the data (changes default tests also).

   b) GIVEN thry/subst specification into empirical model, can estimate complex interactivity, ubiquitous endogeneity, etc. (plus side benefits)…but must give that given…

4. EITM means many different things; all permutations of acronym find proponents

   a) **EITM:** Empirical Implications of Theoretical Models

      (1) *Vision:* Theory ⇒ more, sharper predictions ⇒ better tests, which ∴ inform theory more, …

      (2) *Examples:* Granato, Lupia, Morton, McCarty (1st time)

   b) **TMEI:** Theory-specified Models for Empirical Inference

      (1) *Vision:* Theory directly & fully structures empirical models & relations b/w observations ⇒ specification & (causal) identification of entire empirical models

      (2) *Examples:* Diermeier, Signorino, Mebane, Achen, Sartori, Smith

   c) **TIEM:** Theoretical Implications of Empirical Measures

      (1) *Vision:* Emp. regularities, findings, measures inform theory dev'p.

      (2) *Examples:* Brady, Ericson, McCarty (2nd time)

d) **EMTI:** Empirical Models of Theoretical Intuitions

(1) ***Vision*:** Intuitions derived from theoretical models specify empirical models. I.e., empirical specification to match intuitions, not model. ***Examples:*** Franzese, Kedar, Achen

(a) Note: Runs strongly counter alternative moves stats & econometrics, & related toward experiments, RCT, non-parametric, matching; there, *model-dependence*=4-letter word. Alternative audiences, rhetorical purposes? Convince skeptic some causal effect exists vs. for the convinced/willing, give richer, portable model of how world works?

(b) *Definition*: Tighter integration, deeper interaction, & better communication b/w thry & emp Work

(c) *Quotes:*

"Empirical results can be and will be only as theoretically informing as the empirical specifications from which they derive are theoretically informed."

"Only theoretically informed empirical models, & only insofar & to degree they are theoretically informed, can inform theory."

"The degree to which empirical model theoretically informed equals the potential of its estimation results to inform theory."

(2) ***Point:*** Specification is everything:

(a) Core implication of theories usu. some:

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}) \text{ , or equivalently,}$$

$$E(\mathbf{y}) = f(\mathbf{x}, \boldsymbol{\beta}), \quad \boldsymbol{\varepsilon} \sim g(\boldsymbol{\varepsilon}).$$

(b) **EMTI** stresses that far too little typically drawn from theoretically implied $f(\cdot)$ & $g(\cdot)$. Usually thry used merely to suggest some $\mathbf{x}$ as arguments, linear-additively by default, to regression or likelihood formula deemed appropriate. Then, most-commonly, hypotheses confined simply & solely to some set of $\partial\mathbf{y}/\partial(x \in \mathbf{x}) <$ or > 0 … Instead, specification of $E(\mathbf{y}|\mathbf{x})$ is the key (and estimation of $E(\mathbf{y}|\mathbf{x})$ the goal).