Spatial Econometric Models of Interdependence:

The Spatial Probit Model: Estimation, Interpretation, and Presentation

Essex Summer School, Lecture 8-9
20 August 2009

drawn from the joint work of

Robert J. Franzese, Jr., *The University of Michigan*

Jude C. Hays, *The University of Illinois*

(***Aya Kachi***, *The University of Illinois*)

The Econometric Problem (1)

- **Spatial Qualitative/Categorical/Lmtd-Dep-Var Models:**
  - **S-Probit:** McMillen 1992,1995; Bolduc et. al. 1997; Pinkse & Slade 1998; LeSage 1999, 2000; Beron et al. 2003; Beron & Vijverberg 2004;
  - **S-Logit:** Dubin 1997; Lin 2003; Autant-Bernard 2006; **S-Selection** (*S-Tobit/Heckit*): McMillen 1995, Smith & LeSage 2004, Flores-Lagunes & Schnier 2006; **S-MNP:** McMillen 1995, Bolduc et al. 1997; **S-Event-History**: Phaneuf & Palmquist 2003; **S-Duration**: Hays & Kachi 2008; **Survival w/ Spatial Frailty:** Banerjee et al. 2004, Darmofal 2007; **S-Count:** Bhati 2005, *including ZIP*: Rathbun & Fei 2006
- **The Challenge:**
  - **In General:** Not *n* conditionally independent observations, so (log)like not product (sum) thereof, but 1 *n*-dimensional (log-)like to max
  - **In S-Probit:** Not *n* uni-D cum-std-norms, but 1 *n*-D heterosked-cum-normals
- **Spatial Latent-Variable Models: Estimation Strategies**
  - McMillen 1992: **EM**, but no s.e.'s rho & arb. param. of induced heterosked.
  - McMillen 1995, Bolduc et. al. 1997: **simulated-likelihood** for S-MNP
  - Beron et al.'03, B&V '04: **RIS**-recursive importance sampling-for S-Prob
  - LeSage 1999, 2000: **Bayesian MCMC** by Metropolis-Hastings-within-Gibbs.
  - Fleming 2004: simpler approximations by S-(non)linear probability models
  - Pinkse & Slade's 1998: **two-step GMM** estimator (for S-error-Probit).

The Econometric Problem (2)

- Structural Model: $\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- Reduced Form: $\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \ \mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$
- Measurement Equation: $y_i = 1 \text{ if } y_i^* > 0\ \ ;\ \ 0 \text{ if } y_i^* \leq 0$
- Probability:

$$p(y_i = 1 \mid \mathbf{X}) = p\left(\left[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i + \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\right]_i > 0\right)$$

$$= p\left(u_i < \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i \big/ \sigma_i\right) = \boldsymbol{\Phi}_i\left\{\left[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i \big/ \sigma_i\right\}$$

- Distrib. **u**: $\boldsymbol{\varepsilon} \sim MVN[0, I],\ \ \mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \Rightarrow \mathbf{u} \sim MVN[0, \boldsymbol{\Sigma}], \text{where}$

$$\boldsymbol{\Sigma} = V[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}] = [(\mathbf{I} - \rho\mathbf{W})^{-1}]' V(\boldsymbol{\varepsilon})[(\mathbf{I} - \rho\mathbf{W})^{-1}]$$

$$= [(\mathbf{I} - \rho\mathbf{W})^{-1}]'\mathbf{I}[(\mathbf{I} - \rho\mathbf{W})^{-1}] = [(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$$

  - For Spatial-Error-Probit:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u};\ \ \mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon} \Rightarrow \mathbf{u} \sim \text{MVN}\left(\mathbf{0}, [(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}\right)$$

$$\Rightarrow\ \ p(y_i = 1 \mid \mathbf{x}_i) = p\left(u_i < \mathbf{x}_i\boldsymbol{\beta} / \sigma_i\right)$$

The Econometric Problem (3)

- Comments:
  - Notice: to calc $\hat{\mathbf{p}}$ or $\frac{\Delta\hat{\mathbf{p}}}{\Delta\mathbf{X}}$, etc., same MVN integration
    - No such substantive interpretation yet in lit. that we know.
  - If can order interdep & ensure only antecedent y^* on RHS, then std-ML probit w/ s-lag **Wy** may work
    - We think usu. indefensible subst'ly/thry'ly, but cf. Swank...
  - Having y, not y^*, on RHS may seem subst'ly or thry'ly desirable in some cases, but not gen'ly logically possible:
    - Problem: outcome, y_i, would indirectly (via spatial feedback) determine y_i^*, but then y_i^* directly determines y_i (Heckman 78).
    - Think we can calc such counterfactuals, though: $\frac{\Delta\hat{\mathbf{p}}_i}{\Delta\mathbf{y}}$
  - Notice similar MVN issue w/ time lags; suggests similar strategies (simpler b/c ordered?) may allow model temp dynamics directly instead of nuisance (e.g., BKT splines)

<u>Estimators</u>: Bayesian Gibbs-MH Sampler, Basic Idea

- Monte Carlo (MC): Given likelihood/posterior, can sample to estimate any quantity of interest, including a pdf, e.g.
- Markov Chain (MC)MC:
  - Each draw depends on previous, so need only conditional like./post.
  - Some theorems indicate, under fairly gen'l conditions, sample dist of parameter draws converges to its distribution under true like./post.
- Gibbs Sampler: simplest of MCMC family:
  - Express each parameter like./post. conditional on others.
  - Cycle to draw each conditional on others' start vals or prev. draw
  - After some sufficient "burn-in", all subsequent param-vector draws follow true multivariate likelihood/posterior.
- Metropolis-Hastings: useful when cond'l param-dist non-std
  - Draws from a *seed* or *jump* distribution are accepted or rejected as the next sampled parameters, depending on how they compare to a suitably transformed expression of the target distribution

<u>Estimators</u>: Bayesian Gibbs-MH Sampler (2)

- Bayesian Gibbs-MH (MCMC) Sampler for S-Probit:
  - Likelihood: $L\left(\mathbf{y}^*, \mathbf{W} \mid \rho, \boldsymbol{\beta}, \sigma^2\right) = \frac{1}{2\pi\sigma^{2(n/2)}} \left|\mathbf{I}_n - \rho\mathbf{W}\right| e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}$, with

    $\boldsymbol{\varepsilon} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}$ in s-lag; $\boldsymbol{\varepsilon} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)\left(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)$ in s-err
  - Diffuse Priors => Joint Posterior:

    $$p\left(\rho, \boldsymbol{\beta}, \sigma \mid \mathbf{y}^*, \mathbf{W}\right) \propto \left|\mathbf{I}_n - \rho\mathbf{W}\right| \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}$$
  - Conditional Posteriors:
    - $p\left(\sigma \mid \rho, \boldsymbol{\beta}\right) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}$, so $\sigma^2 \sim \chi_n^2$, which is std, so Gibbs
    - $p\left(\boldsymbol{\beta} \mid \rho, \sigma\right) \sim N\left[\tilde{\boldsymbol{\beta}}, \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\right]$, with $\mathbf{C} = \mathbf{I}_n$, $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_n - \rho\mathbf{W}\right)\mathbf{y}^*$ in s-lag

      or $\mathbf{C} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)$ & $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{y}^*$ in s-err
    - $p\left(\rho \mid \boldsymbol{\beta}, \sigma\right) \propto \left|\mathbf{I}_n - \rho\mathbf{W}\right| \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}$, non-std, so MH.
    - $f\left(z_i \mid \rho, \boldsymbol{\beta}, \sigma\right) \sim N(\hat{y}_i^*, \sigma_i^2)$, left- or right-truncated at 0 as $y_i = 1$ or 0

Integration by Simulation

$$\mathbf{y}^* = \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{v} \quad , \text{ with } \mathbf{v} = \boldsymbol{\Gamma}\mathbf{u}$$

$$\mathbf{v} \sim MVN(0, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = \boldsymbol{\Gamma}'\boldsymbol{\Gamma}$$

Brute Force Method: Sample from the reduced form disturbances and calculate the relative frequency of the observed sequence of 1’s and 0’s. This should tell us whether one set of parameter values is more likely to have generated the data we observe than another set…but the sample space is huge and the associated probabilities extremely small (e.g., Lerman and Manski, 1981, “On the Use of Simulated Frequencies to Approximate Choice Probabilities.”)

Estimators: Freq. Recursive Import. Smplr (RIS) (1)

- Basic Idea:
  - To approx. n-dim. cum. std-norm., $p = \int_{-\infty}^{\mathbf{x}_0} f_n(\mathbf{x})\, d\mathbf{x}$
  - Re-express as a mean by mult & divide by a std dist. truncated to support of desired integral, (*the Imp. dist*): $g_n^c(\mathbf{x})$
    - $\Rightarrow p = \int_{-\infty}^{\mathbf{x}_0} \frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\, g_n^c(\mathbf{x})\, d\mathbf{x}$
  - Gives probability, p, sought as mean: $\hat{p} = E\left[\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\right] \approx \frac{1}{R}\sum_{r=1}^{R} \frac{f_n(\tilde{\mathbf{x}}_r)}{g_n^c(\tilde{\mathbf{x}}_r)}$
- $f_n(\tilde{\mathbf{x}}_r)$ in S-Probit is MV-het cum-norm: $p(\mathbf{u} < \mathbf{v}),\ \mathbf{u} \sim MVN\left(\mathbf{0}, \mathbf{\Sigma}\right)$, with

  $\mathbf{\Sigma} = \left(\mathbf{I} - \rho\mathbf{W}\right)' \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1}$, and $\mathbf{v} = \mathbf{Q}\left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \mathbf{X\beta}$, $\mathbf{Q}$ diagonal: $q_i = 2y_i$-1
  - Imp. dist. is n-dim. MVN eval'd at **v**. (uh-oh; but…☺)
  - V-Cov **u** being pos-def => Cholesky decomp. exists s.t.: $\mathbf{\Sigma}^{-1} = \mathbf{A}'\mathbf{A}$, with $\mathbf{A}$ upper-triangular and $\mathbf{\eta} = \mathbf{Au}$ independent!

    So substituting $\mathbf{u} = \mathbf{A}^{-1}\mathbf{\eta} \equiv \mathbf{B\eta}$ gives: $p\left[\mathbf{B\eta} < \mathbf{v}\right] = p\left[\mathbf{\eta} < \mathbf{B}^{-1}\mathbf{v}\right]$

Estimators: Freq. Recursive Import. Smplr (RIS) (2)

- So need calc. this set of *indep*. cum. std.norms:

$$\Pr\left(\mathbf{B\eta}<\mathbf{v}\right)=\Pr\left(\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & \cdots & b_{1,n} \\ 0 & b_{2,2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{n\text{-}1,n\text{-}1} & b_{n\text{-}1,n} \\ 0 & \cdots & 0 & 0 & b_{n,n} \end{bmatrix}\begin{bmatrix} \eta_1 \\ \vdots \\ \vdots \\ \eta_{n\text{-}1} \\ \eta_n \end{bmatrix}<\begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_{n\text{-}1} \\ v_n \end{bmatrix}\right)$$

- Which is this:

$$\Pr\left(\begin{bmatrix} \sum_{i=1}^{n} b_{1,i}\eta_i \\ \vdots \\ b_{n\text{-}1,n\text{-}1}\eta_{n\text{-}1}+b_{n\text{-}1,n}\eta_n \\ b_{n,n}\eta_n \end{bmatrix}<\begin{bmatrix} v_1 \\ \vdots \\ v_{n\text{-}1} \\ v_n \end{bmatrix}\right)=\Pr\left(\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_{n\text{-}1} \\ \eta_n \end{bmatrix}<\begin{bmatrix} b_{1,1}^{-1}\left(v_1-\sum_{i=2}^{n} b_{1,i}\eta_i\right) \\ \vdots \\ \vdots \\ b_{n\text{-}1,n\text{-}1}^{-1}\left(v_{n\text{-}1}-b_{n\text{-}1,n}\eta_n\right) \\ b_{n,n}^{-1}v_n \end{bmatrix}\right)$$

- Get cutpoints recursively, starting w/ last:
  - Calculate upper bound for truncated-normal dist. of n^{th}
  - Draw from this dist for upper bound for $(n\text{-}1)^{th}$…
  - Since indep., probability of sample (0,1) observed is product of *n* univariate cum std. norms at these bounds
  - Repeat *R* times & avg => RIS est. of the *log-l*ikelihood to max:

Estimators: Freq. Recursive Import. Smplr (RIS) (2)

- Get cutpoints recursively, starting w/ last:
  - Calculate upper bound for truncated-normal dist. of nth
  - Draw from this dist for upper bound for (n-1)th…
  - Since indep., prob of sample (0,1) observed is product of *n* univariate cum std. norms at these bounds…

$$\left.\begin{array}{c} \eta_n < b_{n,n}^{-1} v_n \equiv \upsilon_n \\ \eta_{n\text{-}1} < b_{n\text{-}1,n\text{-}1}^{-1}\left[v_{n\text{-}1} - b_{n\text{-}1,n}\tilde{\eta}_n\right] \equiv \upsilon_{n\text{-}1} \\ \eta_{n\text{-}2} < b_{n\text{-}2,n\text{-}1}^{-1}\left[v_{n\text{-}2} - b_{n\text{-}2,n\text{-}1}\tilde{\eta}_{n\text{-}1} - b_{n\text{-}2,n}\tilde{\eta}_n\right] \equiv \upsilon_{n\text{-}2} \\ \vdots \end{array}\right\} \Rightarrow \eta_j < b_{j,j}^{-1}\left[v_j - \sum_{i=j+1}^{n} b_{j,i}\tilde{\eta}_i\right] \equiv \upsilon_j$$

  - These indep. marginals, so estimated likelihood is:

$$\prod_{j=1}^{n} p_j = \prod_{j=1}^{n} \Phi\left(\upsilon_j\right)$$

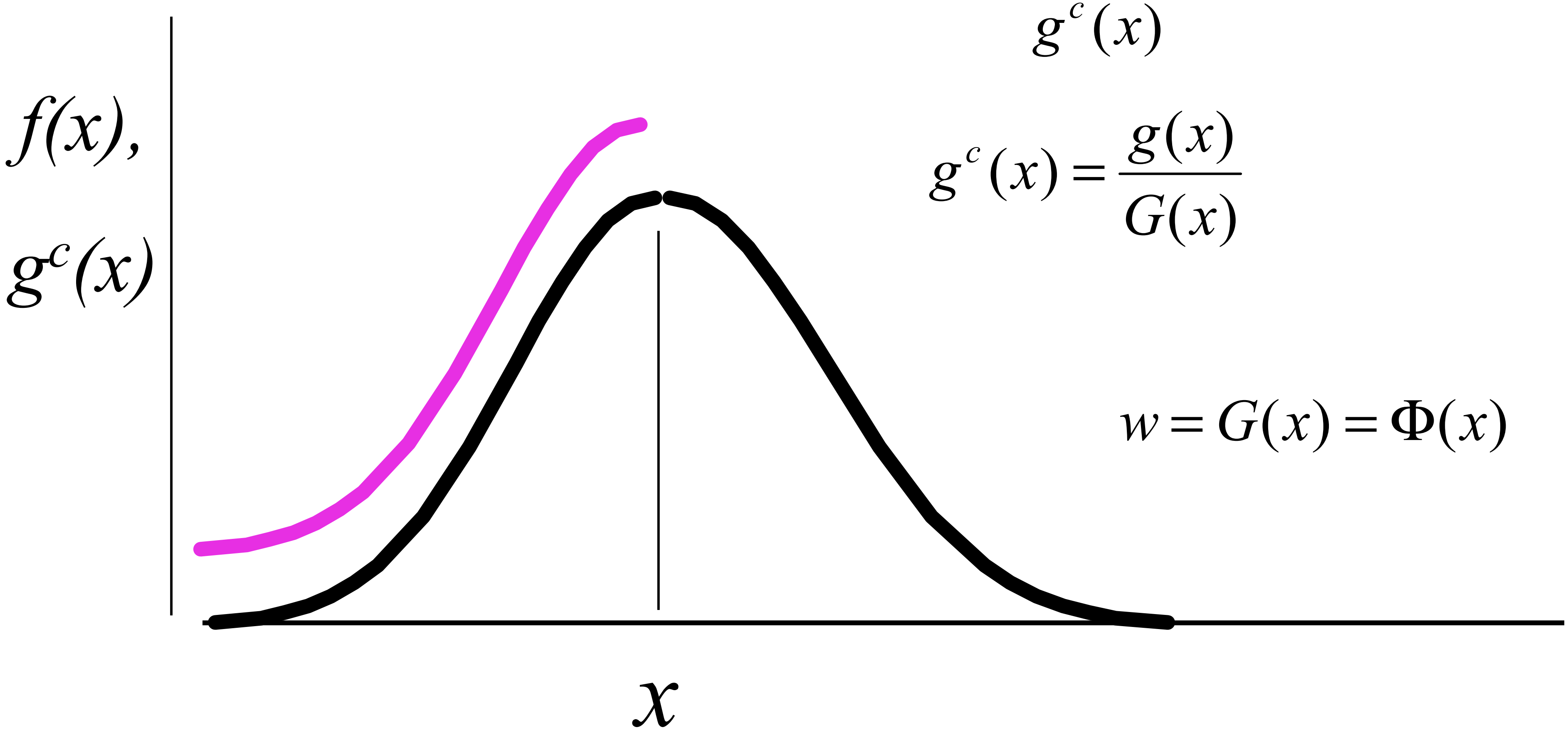
  - Repeat *R* times & avg => RIS est. of *log*-*l*ikelihood to max:

$$\hat{l} = (1/R)\sum_{r=1}^{R}\left[\prod_{j=1}^{n}\Phi\left(\upsilon_{j,r}\right)\right]$$

Integration by Simulation: RIS

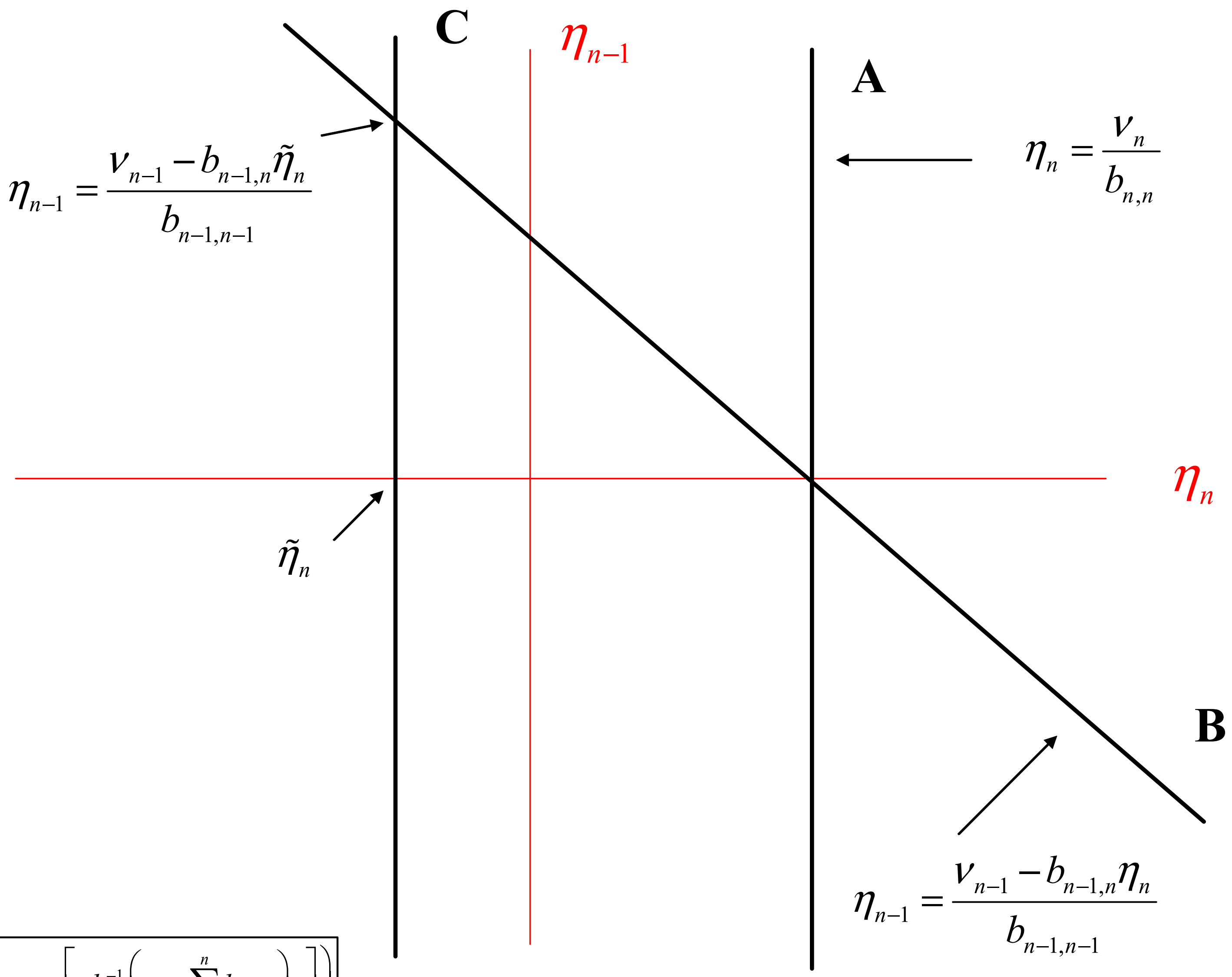
$$w = \frac{f(x)}{g^c(x)}$$

$$g^c(x) = \frac{g(x)}{G(x)}$$



$$w = G(x) = \Phi(x)$$

RIS



$$\Pr\left(\begin{bmatrix} \sum_{i=1}^{n} b_{1,i}\eta_i \\ \vdots \\ b_{n\text{-}1,n\text{-}1}\eta_{n\text{-}1} + b_{n\text{-}1,n}\eta_n \\ b_{n,n}\eta_n \end{bmatrix} < \begin{bmatrix} v_1 \\ \vdots \\ v_{n\text{-}1} \\ v_n \end{bmatrix}\right) = \Pr\left(\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_{n\text{-}1} \\ \eta_n \end{bmatrix} < \begin{bmatrix} b_{1,1}^{-1}\left(v_1 - \sum_{i=2}^{n} b_{1,i}\eta_i\right) \\ \vdots \\ \vdots \\ b_{n\text{-}1,n\text{-}1}^{-1}\left(v_{n\text{-}1} - b_{n\text{-}1,n}\eta_n\right) \\ b_{n,n}^{-1}v_n \end{bmatrix}\right)$$

Evaluating the Estimators (One Quick MC)

- DGF: same **W**, diff. r

$$\mathbf{y}^* = \left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\left(\mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right), \text{ with } \mathbf{x} = \left(\mathbf{I}_n - \theta\mathbf{W}\right)^{-1}\mathbf{z} \;\&\; \mathbf{z}, \boldsymbol{\varepsilon} \sim N(0,1)$$

- Conditions:
  - Row-std contig U.S. 48
  - r=0.5, β=1.0, n={48,144}, θ={0.0, 0.5}
- You can't see this, but…
- Rel'ly poor bias perf. BG, though better RMSE.
- Even std ML w/ **Wy** seems dom, but this b/c 2 counteracting biases, meas./spec. err & simult.
  - Simult incr in r , meas-err decr/flat in n, so over- to under-est. (Seems true.)
- B&V '04 do MC like exp 2 & find r=-18%, β=+10%, so RIS seems dom on both criteria, but ***SLOW***. (MCMC only *SLOW*)

Table 1: Simulation Results

	ML with Wy		ML with Wy*		Bayesian MCMC	
	β	ρ	β	ρ	β	ρ
Experiment #1: n=48, θ=0.0						
Mean Coefficient Estimate	1.02	0.32	1.13	0.74	1.23	0.30
Root Mean-Squared Error	0.33	0.71	0.43	0.43	0.36	0.26
Actual Std Dev of Estimates	0.33	0.69	0.41	0.36	0.28	0.16
Mean of Reported Std Err	0.30	0.41	0.35	0.30	0.42	0.21
Experiment #2: n=48, θ=0.5						
Mean Coefficient Estimate	1.22	0.35	1.13	0.69	1.21	0.28
Root Mean-Squared Error	0.60	0.77	0.62	0.38	0.32	0.26
Actual Std Dev of Estimates	0.56	0.76	0.61	0.33	0.24	0.14
Mean of Reported Std Err	0.36	0.46	0.42	0.29	0.39	0.20
Experiment #3: n=144, θ=0.0						
Mean Coefficient Estimate	0.94	0.42	1.01	0.68	1.14	0.34
Root Mean-Squared Error	0.18	0.28	0.19	0.24	0.21	0.19
Actual Std Dev of Estimates	0.17	0.27	0.19	0.16	0.15	0.10
Mean of Reported Std Err	0.16	0.22	0.18	0.15	0.22	0.12
Experiment #4: n=144, θ=0.5						
Mean Coefficient Estimate	1.08	0.48	0.97	0.64	1.13	0.32
Root Mean-Squared Error	0.21	0.29	0.21	0.21	0.19	0.20
Actual Std Dev of Estimates	0.19	0.29	0.21	0.16	0.14	0.09
Mean of Reported Std Err	0.18	0.23	0.20	0.15	0.21	0.12

Evaluating the Estimators (A Second Quick MC)

Table 1. Spatial Probit Monte Carlo Results (100 Samples)

Parameter	Result	B&V(RIS)	B&V(LPM)	Hays(RIS)	Hays(Wy)
	mean	3.29 / .85*	.89*	.95	.91
	s.d.	1.08	.17	.27	.26
$B\&V: \hat{\beta}=3$ $Hays: \hat{\beta}=1$	rmse	1.12	N.A.	.28	.28
	mean $\widehat{s.e.}$	N.A.	N.A.	.28	.27
	overconfidence	N.A.	N.A.	.97	.96
	mean	.41	.25	.35	.40
	s.d.	.22	.16	.20	.36
$\hat{\rho}=.5$	rmse	.24	.30	.25	.37
	mean $\widehat{s.e.}$	N.A.	N.A.	.20	.36
	overconfidence	N.A.	N.A.	1.00	.99

Notes: R=2000 vs. R=100, N=50 vs. N=48 ; * marginal effects

Calculating & Presenting *Effects* (1)

- If confine discussion to ***y****, then as prev. F&H:

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= \begin{bmatrix} 1 & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,n} \\ -\rho w_{2,1} & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & -\rho w_{(n\text{-}1),n} \\ -\rho w_{n,1} & \cdots & \cdots & -\rho w_{n,(n\text{-}1)} & 1 \end{bmatrix}^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \equiv \mathbf{S}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)$$

- And s.e.'s/c.i.'s by delta method as:

$$\hat{V}\left(\hat{\mathbf{s}}_i\hat{\beta}_k\right) \approx \left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\hat{V}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]', \text{ with } \hat{\boldsymbol{\theta}} = \begin{bmatrix}\hat{\rho} \\ \hat{\beta}_k\end{bmatrix} \;\&\; \left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right] = \begin{bmatrix}\dfrac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \hat{\mathbf{s}}_i\end{bmatrix}$$

- Or simulate (parametric bootstrap); reasons to think that would be better, actually…

Calculating & Presenting *Effects* (2)

- But we (should) want discuss:

$$\frac{\Delta \mathbf{p}}{\Delta \mathbf{X}} = \boldsymbol{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_1\boldsymbol{\beta}\right]\odot\left[\left\{\sigma_i^{-1}\right\}\right]\right)-\boldsymbol{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_0\boldsymbol{\beta}\right]\odot\left[\left\{\sigma_i^{-1}\right\}\right]\right)$$

  - Note the complications:
    - Given probit, must know $\mathbf{x}_i$; given s-lag interdep., must know $\mathbf{X}$ (!). (S-err only $\mathbf{x}_i$)
    - Given interdep., calc more meaningful of these $\hat{\mathbf{p}}$, the conditional ones, will req. MVN *cdf*. (S-lag or S-err.)
- Marginal vs. Conditional Counterfactual $\hat{\mathbf{p}}$

Can we use the GHK simulator?

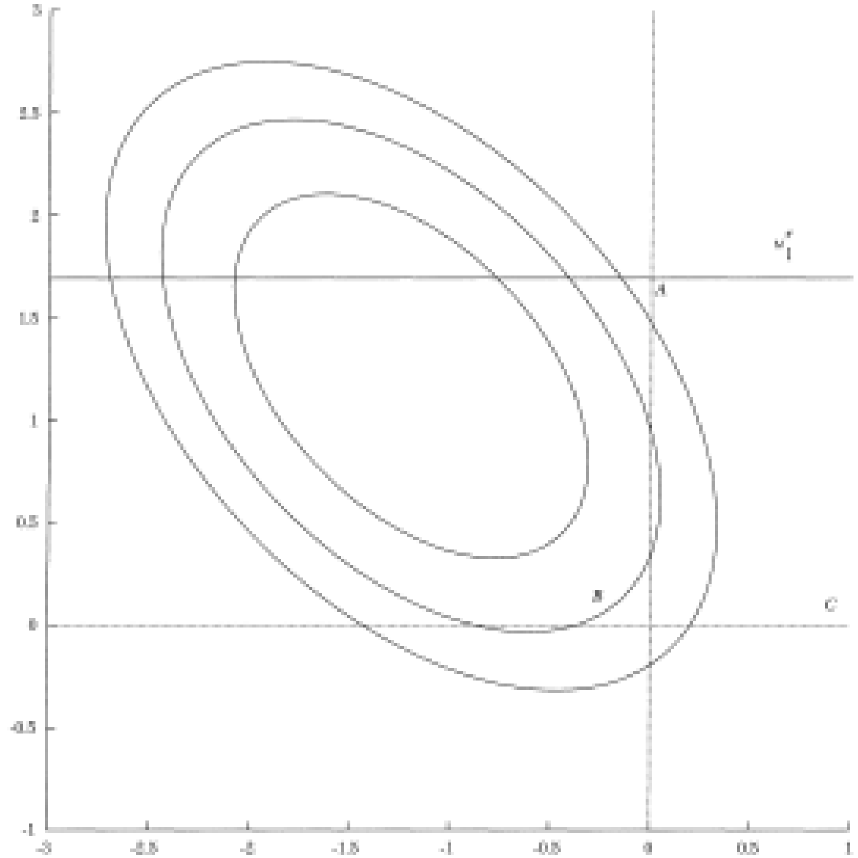


Figure 3. GHK Algorithm

Source: Stern (1997, 2026)

<u>Example Estimation App.</u>: US State CHIP Premia

	Probit-ML	Probit-MCMC	Spatial-Lag Probit (Gibbs)	Spatial-Error Probit (Gibbs)	Spatial-Lag Probit (RIS)
Constant	-4.978	-5.163	-5.606	-5.531	-5.186
	(6.260)	(6.292)	(10.159)	(7.337)	(5.944)
Poverty Rate	-.244	-.265**	-.374**	-.243*	-.171
	(.153)	(.156)	(.231)	(.157)	(.125)
Retail Wage	.004	.004*	.006*	.004*	.004*
	(.003)	(.003)	(.004)	(.003)	(.002)
Government Ideology	.011	.011	.014	.014	-.004
	(.013)	(.013)	(.020)	(.014)	(.116)
Inter-party Competition	2.174	2.108	1.473	2.636	1.27
	(3.388)	(3.478)	(6.134)	(3.794)	(31.94)
Tax Effort	-.014	-.014	-.020	-.017	.005
	(.019)	(.019)	(.034)	(.021)	(.017)
Federal Share	.045	.048	.065	.043	.041
	(.063)	(.064)	(.095)	(.066)	(.056)
Spatial lag or error-lag	.079	.102	.200***	.297***	.243
	(.798)	(.815)	(.148)	(.196)	(.252)
Pseudo-R^2	.222	.220	.607	.574	NA
Observations	48	48	48	48	48

Notes: 1. Informative U(0,1) prior on *rho* helps. We've qualms.

2. Difference in Bayesian vs. frequentist significance also.

3. Note measurement/specification-error seem to have dom'd here for ML.

<u>Example Estimation App.</u>: U.S. State Leg. Term Limits

<u>Table 3</u>: Adoption of Term Limits for State Legislators, Estimation Results

	Probit-ML	Probit-MCMC	Spatial-Lag Probit	Spatial-Error Probit	Probit-RIS
Constant	-.539	-.909	-.598	-.411	.313
	(2.579)	(1.814)	(3.080)	(2.533)	(2.223)
I&R	2.320***	1.806***	3.257***	2.650***	2.336***
	(.581)	(.481)	(.917)	(.627)	(.620)
Clinton	.273	.131	.147	.056	.304
	(.542)	(.476)	(.769)	(.606)	(.518)
Tax Effort	-.178	-.068	-.176	-.146	-.214
	(.273)	(.183)	(.321)	(.262)	(.247)
Spatial lag or error-lag	.926	.634	.144	.018	.416**
	(.801)	(.687)	(.207)	(.279)	(.194)
Pseudo-R^2	.480	.458	.833	.803	—
Log-Likelihood	-17.093	—	—	—	-16.261
Observations	48	48	48	48	48

Notes: 1. Informative U(0,1) prior on *rho* helps. We've qualms.

2. Difference in Bayesian vs. frequentist significance also.

3. Note simult bias seems to dominate for std-ML or MCMC w/ s-lag **Wy**.

Example Estimation App.:

Table 3. Spatial Models of WWI Participation and Entry Timing

	(1)	(2)	(3)	(4)	(5)
Constant	-.806	8.19***	9.87***	9.81***	11.58***
	(.611)	(2.21)	(2.19)	(2.14)	(2.54)
Shape Parameter		3.33***	3.42***	3.40***	3.32***
		(.650)	(.685)	(.676)	(.656)
Contiguity Spatial Lag	.482**	.289			
	(.217)	(.180)			
Targeted Alliance Spatial Lag			.180		
			(1.32)		
Rivalry Spatial Lag				.288	
				(.491)	
Territorial Dispute Spatial Lag					-2.42**
					(1.21)
National Capabilities	18.48***	-41.45***	-38.03**	-39.68**	-38.18**
	(6.61)	(15.62)	(17.98)	(15.69)	(15.37)
Democracy	-.055	.045	.081	.09	.040
	(.137)	(.118)	(.129)	(.116)	(.119)
Trade	-.156*	-.088	-.178	-.149	-.469
	(.088)	(.331)	(.336)	(.338)	(.370)
Europe	1.54**	-4.52***	-5.48***	-5.47***	-4.51***
	(.648)	(1.65)	(1.77)	(1.65)	(1.66)
Model	Probit	Weibull	Weibull	Weibull	Weibull
Observations	44	44	44	44	44
Log-Likelihood	-14.616	-48.151	-49.332	-49.190	-47.079

Notes: In the probit case, the dependent variable reflects participation in WWI (0=No, 1=Yes). For the duration models, the dependent variable is the number of months before entering WWI. Of the 44 sample countries, 15 enter the War. All the spatial weights matrices are row-standardized. *National Capabilities* are the COW CINC index scores. *Democracy* is Polity measures of regime type. *Trade* is the value of total trade in current US dollars (Source: Barbieri 2002). Europe is a dummy variable that takes a value of 1 for countries located on the continent, including the United Kingdom. Parentheses contain standard error estimates. ***significant at 1%; **significant at 5%; *significant at 10%.

Back to CF Probabilities

Can we condition the probability for unit *i* on the outcome for unit *j? GHK probabilities are problematic, but what if we try the brute force method?*

$$\mathbf{y}^* = \mathbf{\Gamma X \beta} + \mathbf{v} \ \ , \ \text{with } \mathbf{v} = \mathbf{\Gamma u}$$

$$\mathbf{v} \sim MVN(0, \mathbf{\Omega}) \qquad \mathbf{\Omega} = \mathbf{\Gamma}'\mathbf{\Gamma}$$

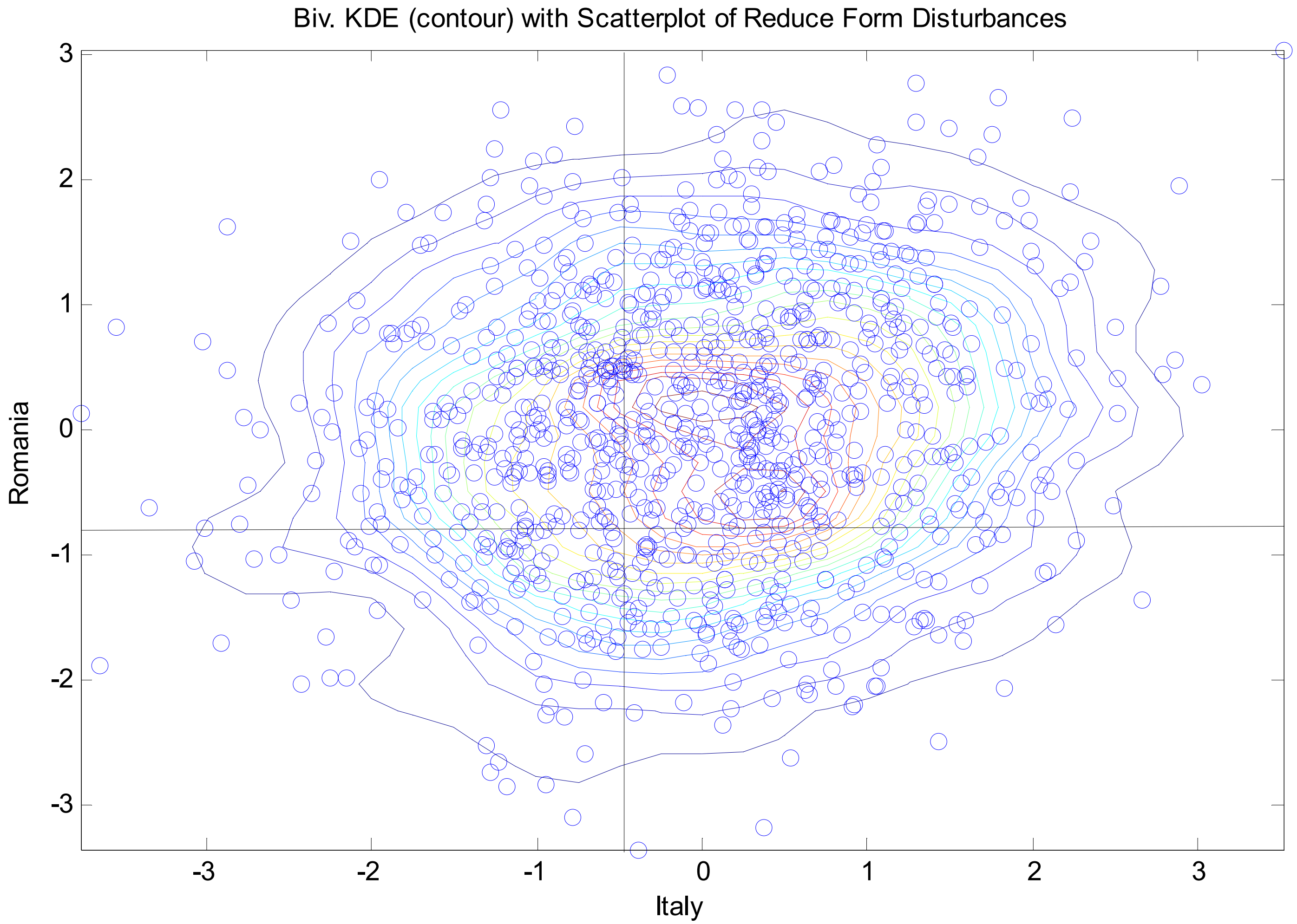
Sample from the reduced form disturbances. Split the sample based on *j*. Compute the relative frequencies of 0's and 1 conditional on *j*. *This strategy may not work well for estimating parameters, but should be able to estimate counterfactual probabilities.*

What about CF Probabilities?

Can we condition the probability for unit *i* on the outcome for unit *j*?

How did America's decision to participate affect the participation of other states? To what extent did Italy's decision to enter the war in mid-1915 affect the probability that Bulgaria and Romania would be drawn into the conflict before the fighting stopped?

How did Italy's Participation Affect Romania?



$$\mathbf{y^*} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{X\beta} + (\mathbf{I} - \mathbf{A})^{-1}\mathbf{u}$$

$$= \mathbf{\Gamma X\beta} + \mathbf{\Gamma u}$$

$$= \mathbf{\Gamma X\beta} + \mathbf{v}$$