Endogeneity & Accounting

part III

Doug Schroeder
Ohio State University

Purdue University workshop
October 18-19, 2007

*Outline – **Bayesian analysis of self-selection***

I. Bayesian simulation & McMC (emphasis on Gibbs sampler)

II. Gibbs sampler applied to probit – Albert and Chib

III. Gibbs sampler applied to self-selection

IV. Some archival results

V. Concluding remarks

Bayesian inference

In Bayesian analysis, we typically generate simulations based on the marginal posterior distribution (product of likelihood and prior).

$p(\theta|y) = p(y|\theta)p(\theta)/p(y)$

but the denominator is a normalizing constant so we usually focus on the kernel

$p(\theta|y) \propto p(y|\theta)p(\theta)$

The Gibbs sampler

If we cannot derive $p(\theta|y)$ in closed form (if it does not have a standard probability distribution), we can utilize full conditional posterior distributions to draw dependent samples for the parameters of interest via McMC simulation.

The Gibbs sampler (cont.)

Full conditional posterior distributions:

$p(\theta_1| \theta_2, \theta_3, y)$ draws are made for θ_1 conditional on starting values for θ_2, θ_3.

$p(\theta_2| \theta_1, \theta_3, y)$ then, θ_2 is drawn conditional on the θ_1 draw and starting value for θ_3.

$p(\theta_3| \theta_1, \theta_2, y)$ next, θ_3 is drawn conditional on the draws for θ_1 and θ_2.

The Gibbs sampler (cont.)

Sampling continues for a large number of draws with parameters updated each iteration by the most recent draw. A post-convergence sample is employed for inferences.

The Gibbs sampler (cont.)

– The samples are dependent.

– Not all samples will be from the marginal posterior; only after a finite (but unknown) number of iterations are draws from the marginal posterior.

– Note, in general, $p(\theta_1, \theta_2|y) \neq p(\theta_1|\theta_2, y)\, p(\theta_2|\theta_1, y)$.

Convergence checks include trace plots, burn-in iterations, and other convergence diagnostics.

Data augmentation

Albert and Chib's Gibbs sampler Bayes probit

The challenge with discrete choice models (like probit) is that the latent utility is unobservable

rather the analyst observes only discrete (usually binary) choices

Albert & Chib employ Bayesian data augmentation to "supply" the latent variable. Hence, parameters of a probit model are estimated via normal Bayesian regression

$$U_D = W\theta - V$$

Albert and Chib's Gibbs sampler Bayes probit (cont.)

$p(\theta\ |D, W, U_D) \sim N(b_1, (Q^{-1} + W^TW)^{-1})$

where $b_1 = (Q^{-1} + W^TW)^{-1}(Q^{-1} b_0 + W^TWb)$,

$b = (W^TW)^{-1} W^TU_D$,

b_0 = prior means for θ and

$Q = (W_0^TW_0)^{-1}$ is the prior on the covariance.*

*details next slide

Albert and Chib's Gibbs sampler Bayes probit (cont.)

*In general, Bayesian regression works as if we have data from the prior period $\{y_0, X_0\}$ as well as from the sample period $\{y, X\}$ from which β is estimated.

Applying OLS to $\begin{bmatrix} y_0 \\ y \end{bmatrix} = \begin{bmatrix} X_0 \\ X \end{bmatrix}\beta + \begin{bmatrix} \varepsilon_0 \\ \varepsilon \end{bmatrix}$

yields $b_1 = (X_0^TX_0 + X^TX)^{-1}(X_0^Ty_0 + X^Ty)$

since $Q^{-1} = (X_0^TX_0)$, $X_0^Ty_0 = Q^{-1} b_0$, and $X^Ty = X^TXb$,

$b_1 = (Q^{-1} + X^TX)^{-1}(Q^{-1} b_0 + X^TXb)$.

Albert and Chib's Gibbs sampler Bayes probit (cont.)

$p(U_D\ |D = 1, W, \theta) \sim N(W\theta, I|U_D > 0)$ or $TN_{(0,\infty)}(W\theta, I)$

$p(U_D\ |D = 0, W, \theta) \sim N(W\theta, I|U_D \leq 0)$ or $TN_{(-\infty,0)}(W\theta, I)$

random draws from a truncated normal (truncated below for the first and truncated above for the second) produce U_D.

Iterative draws for $(U_D\ |D, W, \theta)$ and $(\theta\ |D, W, U_D)$ form the Gibbs sampler. Post-convergence draws for $(\theta\ |D, W, U_D)$ supply interval estimates of θ.

Self-selection

DGP: $y_j = X\beta_j + \nu_j \qquad j = 0, 1$ (outcome equations)

$U_D = W\theta - V_D$ (selection equation)

$D = 1 \quad U_D > 0$

$D = 0 \quad$ otherwise

$y = D\, y_1 + (1-D)\, y_0$ (observable response)

$\Sigma = Var[V_D, \nu_1, \nu_0]$

Utilize usual IV exclusion restrictions as W contains instruments Z

In the selection setting, there are effectively three sources of missing data:

- the latent utility index U_D
- the counterfactual $(y_1|D = 0)$
- the counterfactual $(y_0|D = 1)$

If these were observable identification of treatment effects $(y_1 - y_0)$ would be straightforward.

Li, Poirier, and Tobias' Gibbs sampler for Bayes selection analysis

For the selection setting, Bayesian inference proceeds analogously to Bayesian probit with data augmentation.

One complication is that $\Sigma = Var[V, \nu_1, \nu_0]$ is unknown and since the counterfactuals are unobserved, $Corr[\nu_1, \nu_0] = \rho_{10}$ is unidentified.

LPT propose using the positive definiteness of Σ to bound ρ_{10}.

LPT's Gibbs sampler Bayes selection analysis (cont.)

From the positivity of the determinant (eigenvalues) of Σ,

$$\rho_{1D}\rho_{0D} - [(1-\rho_{1D}^2)(1-\rho_{0D}^2)]^{1/2} \leq \rho_{10} \leq \rho_{1D}\rho_{0D} + [(1-\rho_{1D}^2)(1-\rho_{0D}^2)]^{1/2}$$

The more pressing is the endogeneity problem, the tighter are the bounds.

This allows learning about ρ_{10} and utilization of the full covariance.

Hence, distributions of treatment effects rather than simply their means can be identified (at least bounded).

LPT's Gibbs sampler Bayes selection analysis (cont.)

Define the complete or augmented data as

$$r_i^* = [U_{Di}, D_i\, y_i + (1-D_i)\, y_i^{miss}, D_i\, y_i^{miss} + (1-D_i)\, y_i]^T$$

$$H_i = \begin{bmatrix} W_i & 0 & 0 \\ 0 & X_i & 0 \\ 0 & 0 & X_i \end{bmatrix} \text{ and}$$

$$\beta = [\theta, \beta_1, \beta_0]^T.$$

LPT's Gibbs sampler Bayes selection analysis (cont.)

Let Γ_{-x} denote all parameters other than x.

Full conditional posteriors for the augmented data:

$$y_i^{miss} \mid \Gamma_{-y_i^{miss}}, Data \sim N((1-D_i)\mu_{1i} + D_i\mu_{0i}, (1-D_i)\omega_{1i} + D_i\omega_{0i})$$

where

$$\mu_{1i} = X_i\beta_1 + \left[\frac{\sigma_0^2\sigma_{1D} - \sigma_{10}\sigma_{0D}}{\sigma_0^2 - \sigma_{0D}^2}\right](U_{Di} - W_i\theta) + \left[\frac{\sigma_{10} - \sigma_{0D}\sigma_{1D}}{\sigma_0^2 - \sigma_{0D}^2}\right](y_i - X_i\beta_0)$$

$$\mu_{0i} = X_i\beta_0 + \left[\frac{\sigma_1^2\sigma_{0D} - \sigma_{10}\sigma_{1D}}{\sigma_1^2 - \sigma_{1D}^2}\right](U_{Di} - W_i\theta) + \left[\frac{\sigma_{10} - \sigma_{0D}\sigma_{1D}}{\sigma_1^2 - \sigma_{1D}^2}\right](y_i - X_i\beta_1)$$

LPT's Gibbs sampler Bayes selection analysis

Full cond. posteriors for the augmented data (cont.):

$$\omega_{1i} = \sigma_1^2 - \frac{\sigma_{1D}^2\sigma_0^2 - 2\sigma_{10}\sigma_{0D}\sigma_{1D} + \sigma_{10}^2}{\sigma_0^2 - \sigma_{0D}^2}$$

$$\omega_{0i} = \sigma_0^2 - \frac{\sigma_{0D}^2\sigma_1^2 - 2\sigma_{10}\sigma_{0D}\sigma_{1D} + \sigma_{10}^2}{\sigma_1^2 - \sigma_{1D}^2}$$

LPT's Gibbs sampler Bayes selection analysis

Full cond. posteriors for the augmented data (cont.):

$U_{Di} \mid \Gamma_{-U_{Di}}, Data \sim TN_{(0,\infty)}(\mu_{ui}, \omega_D) \quad if\ D_i = 1$

$\qquad TN_{(-\infty,0)}(\mu_{ui}, \omega_D) \quad if\ D_i = 0$

$$\mu_{ui} = W_i\theta + (D_i\, y_i + (1-D_i)\, y_i^{miss} - X_i\beta_1)\left[\frac{\sigma_0^2\sigma_{1D} - \sigma_{10}\sigma_{0D}}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2}\right]$$

$$+ (D_i\, y_i^{miss} + (1-D_i)\, y_i - X_i\beta_0)\left[\frac{\sigma_1^2\sigma_{0D} - \sigma_{10}\sigma_{1D}}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2}\right]$$

$$\omega_D = 1 - \frac{\sigma_{1D}^2\sigma_0^2 - 2\sigma_{10}\sigma_{0D}\sigma_{1D} + \sigma_1^2\sigma_{0D}^2}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2}$$

LPT's Gibbs sampler Bayes selection analysis

Full conditional posteriors for the parameters:

$\beta \mid \Gamma_{-\beta}, Data \sim N(\mu_\beta, \omega_\beta)$ with prior $p(\beta) \sim N(\beta_0, V_\beta)$

where

$\mu_\beta = [H^T(\Sigma^{-1} \otimes I_n)H + V_\beta^{-1}]^{-1}\, [H^T(\Sigma^{-1} \otimes I_n)r + V_\beta^{-1}\beta_0]$

and

$\omega_\beta = [H^T(\Sigma^{-1} \otimes I_n)H + V_\beta^{-1}]^{-1}$.

(This is the SUR generalization of Bayesian regression.)

LPT's Gibbs sampler Bayes selection analysis

Full conditional posteriors for the parameters (cont.):

$\Sigma \mid \Gamma_{-\Sigma}, Data = G^{-1}$,

$G \sim Wishart(n+\varrho,\ [S + \varrho R])$

with prior $p(G) = Wishart(\varrho, \varrho R)$

where $S = \sum_{i=1,n} (r_i - H_i\beta)(r_i - H_i\beta)^T$ and σ_D^2 is normalized to one.

LPT's Gibbs sampler Bayes selection analysis

Full conditional posteriors for the parameters (cont.):

Normalization to one creates a slight complication as the full conditional is no longer inverse-Wishart.

Nobile (2000) provides a convenient algorithm for random Wishart (multivariate χ^2) draws with a restricted element.

LPT's Gibbs sampler Bayes selection analysis

Nobile's normalized inverted-Wishart algorithm (cont.):

1. Exchange rows and columns 1 and 3 in $S + \varrho R$, call this matrix V.
2. Find L such that $V = (L^{-1})^T L^{-1}$.
3. Construct a lower triangular matrix A with
   - a. a_{ii} equal to the square root of χ^2 random variates, i=1, 2
   - b. $a_{33} = 1/l_{33}$,
   - c. a_{ij} equal to $N(0,1)$ random variates, $i > j$.

LPT's Gibbs sampler Bayes selection analysis

Nobile's normalized inverted-Wishart algorithm (cont.):

4. Set $V' = (L^{-1})^T (A^{-1})^T A^{-1} L^{-1}$.
5. Exchange rows and columns 1 and 3 in V' and denote this draw Σ.

LPT's Gibbs sampler Bayes selection analysis

Robustness:

To explore *robustness* a *two component mixture* of normals is considered.

The contribution of one individual to the likelihood is

$p(r_i | \Gamma) = \pi_1 \phi(r_i; H_i\beta^1, \Sigma^1) + \pi_2 \phi(r_i; H_i\beta^2, \Sigma^2)$

where each component has its own parameter vector β^j and covariance matrix Σ^j, and $\pi_1 + \pi_2 = 1$.

LPT's Gibbs sampler Bayes selection analysis

Full conditional posteriors for the components are

$c_i | \Gamma_{-c}, Data \sim Multinomial(1, p^1, p^2)$*

$$\text{where } p^j = \frac{\pi_j |\Sigma|^{-1/2} \exp\left[-0.5\left(r_i - H_i\beta^j\right)^T \left(\Sigma^j\right)^{-1}\left(r_i - H_i\beta^j\right)\right]}{\sum_{j=1}^{2} \pi_j |\Sigma|^{-1/2} \exp\left[-0.5\left(r_i - H_i\beta^j\right)^T \left(\Sigma^j\right)^{-1}\left(r_i - H_i\beta^j\right)\right]}$$

*binomial suffices for 2 component mixture

LPT's Gibbs sampler Bayes selection analysis

Full conditional posteriors for the component probabilities are

$\pi_i | \Gamma_{-\pi}, Data \sim Dirichlet(n_1 + \alpha_1, n_2 + \alpha_2)$ with prior hyperparameter α_j

where $n_j = \Sigma_{i=1,n}\, c_{ji}$

Predictive distribution for treatment effects

The Gibbs sampler for the selection problem can be used to generate treatment effect predictive distributions $(y_1^f - y_0^f)$ conditional on X^f *and* W^f utilizing the post-convergence parameter draws.

Predictive distribution for treatment effects

Treatment effect:

$p(y_1^f - y_0^f | X^f) \sim N(X^f(\beta_1 - \beta_0, \sqrt{\gamma_2})$

where $\gamma_2 = Var[y_1^f - y_0^f | X] = \sigma_1^2 + \sigma_0^2 - 2\,\sigma_{10}$.

Predictive distribution for treatment effects (cont.)

Treatment effect on treated:

$p(y_1^f - y_0^f | X^f, D(W^f) = 1)$

$= p(y_1^f - y_0^f | X^f) p(D(W^f) = 1 | y_1^f - y_0^f, X^f) / p(D(W^f) = 1)$

where $p(D(W^f) = 1 | y_1^f - y_0^f, X^f)$

$\sim \Phi(\{W^f\theta + \gamma_1/\gamma_2[(y_1^f - y_0^f) - X^f(\beta_1 - \beta_0)]\} / \sqrt{1 - \gamma_1^2/\gamma_2}\ )$.

$\gamma_1 = Cov[y_1^f - y_0^f, U_D | X, Z] = \sigma_{1D} - \sigma_{0D}$.

Predictive distribution for treatment effects (cont.)

Treatment effect on untreated:

$p(y_1^f - y_0^f \mid X^f, D(W^f) = 0)$

$= p(y_1^f - y_0^f \mid X^f)[1 - p(D(W^f)=1 \mid y_1^f - y_0^f, X^f)]/[1 - p(D(W^f)=1)]$

where $p(D(W^f) = 1 \mid y_1^f - y_0^f, X^f)$

$$\sim \Phi(\{W^f\theta + \gamma_1/\gamma_2[(y_1^f - y_0^f) - X^f(\beta_1 - \beta_0)]\}/\sqrt{1 - \gamma_1^2/\gamma_2}\ ).$$

$\gamma_1 = Cov[y_1^f - y_0^f, U_D \mid X, Z] = \sigma_{1D} - \sigma_{0D}.$

Predictive distribution for treatment effects (cont.)

Local treatment effect:

$p(y_1^f - y_0^f \mid X^f, D(W'^f) = 1, D(W^f) = 0)$

$= p(y_1^f - y_0^f \mid X^f) / [p(D(W'^f) = 1) - p(D(W^f) = 1)]$

$* [p(D(W'^f) = 1 \mid y_1^f - y_0^f, X^f) - p(D(W^f) = 1 \mid y_1^f - y_0^f, X^f)]$

Predictive distribution for treatment effects (cont.)

The foregoing discussion focuses on identifying *conditional* (on the parameters η) predictive distributions.

"Rao-Blackwellization" is utilized to identify *unconditional* predictive distributions.

That is, density ordinates are averaged over parameter draws

$p(y_1^f - y_0^f \mid X^f) = 1/m \sum_{i=1,m} p(y_1^f - y_0^f \mid X^f, \eta = \eta_i)$

where η_i is the ith post-convergence parameter draw of *m*.

Predictive distribution for treatment effects (cont.)

For the "robust" two-component mixture of normals selection analysis, the conditional predictive distributions are the same as above except that we condition on the component and utilize the parameter estimates associated with the component.

The predictive distribution is then based on a probability-weighted average of the components.

Some results based on Hogan's IPO auditor choice data

Hogan's IPO auditor choice data results (cont.)

Outcome equation:

$y_j = X\beta_j + \nu_j$ $j = 0$ (non-big six auditor), 1 (big six auditor)

outcome = first day underpricing (offer value - closing value) + audit fee (all scaled by closing value)

regressors (*X*):

- PREST1 = mid-level (Carter & Manaster 3 to 7.5) underwriter prestige (1); otherwise (0)
- PREST2 = high (Carter & Manaster 8 to 9) underwriter prestige (1); otherwise (0)
- LNASSETS = log of book value of assets at IPO date
- LNOFFVAL = log of offering value at IPO date
- INDRISK = high litigation risk industry (1); otherwise (0)
- DEVEL = new startup/development stage (1); otherwise (0)

Hogan's IPO auditor choice data results (cont.)

selection equation:

$U_D = W\theta - V_D$ $\quad D = 0$ (non-big six auditor)

1 (big six auditor)

regressors (W):

X from outcome equations and

instrument (excluded from X)

UNITTYPE = unit offering (warrants bundled with stock) (1); otherwise (0)

Comparative treatment effect (preliminary) results

are based on

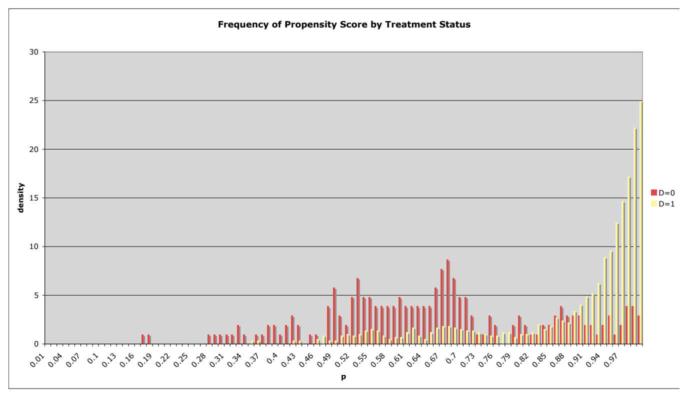
1. Marginal treatment effects via local IV (LIV*)
2. Bayesian data augmentation**
   a. Normal likelihood ($Bayes_1$)
   b. Two component mixture of Normals likelihood ($Bayes_2$)
3. Heckman's two stage normal control function (H2SCF)
4. Dummy variable ordinary least squares (OLS)

*means, standard deviations, and treatment effects based on 50 bootstraps

**treatment effects based on predictive distributions from 1000 post-convergence parameter draws

Propensity score

Regressor	Probit	$Bayes_1$	$Bayes_2$ .79	$Bayes_2$ .21
Intercept	-8.05844 (1.58512)	-8.06864 (1.749890)	-17.2598 (3.07934)	-3.60912 (2.58480)
Prestige 1	0.48329 (0.208058)	0.474323 (0.212340)	0.81756 (0.491384)	1.132831 (0.399137)
Prestige 2	0.53743 (0.25166)	0.501843 (0.267338)	4.47569 (1.464936)	1.07948 (0.497476)
LnAssets	-0.06782 (0.04355)	-0.068087 (0.041571)	0.077871 (0.064066)	-0.22230 (0.082332)
LnOffVal	0.60667 (0.109399)	0.608330 (0.119232)	1.07975 (0.19694)	0.406418 (0.178428)
IndRisk	0.27670 (0.18928)	0.278746 (0.193704)	3.14212 (0.60525)	0.80363 (0.34132)
Devel	0.16718 (0.26063)	0.112108 (0.271130)	-0.068799 (0.208992)	0.0023189 (0.40310)
UnitType	-0.19267 (0.18510)	-0.186441 (0.189264)	-0.49394 (0.353682)	3.90505 (1.53525)



based on probit model

Parameter estimates for outcome equations

D = 1

Regressor	LIV*	$Bayes_1$	$Bayes_2$ .79	$Bayes_2$ .21	H2SCF	OLS
Intercept	na	0.315459 (0.126053)	0.158621 (0.121249)	0.335957 (0.34982)	0.1877620 (0.201008)	0.332583 (0.132699)
Prestige 1	na	-0.010387 (0.018162)	0.006311 (0.017170)	-0.057564 (0.055179)	0.0003197 (0.022231)	-0.012574 (0.019651)
Prestige 2	na	-0.041112 (0.020655)	-0.026767 (0.020153)	-0.071901 (0.069009)	-0.0312348 (0.022505)	-0.042236 (0.022051)
LnAssets	0.0041356 (0.040744)	-0.003427 (0.003690)	0.002754 (0.003366)	-0.040011 (0.009918)	-0.0041720 (0.003315)	-0.003458 (0.003800)
LnOffVal	0.37861 (0.30949)	-0.008374 (0.008783)	-0.006733 (0.0083157)	0.0297017 (0.024669)	-0.0012147 (0.012059)	-0.009199 (0.009215)
IndRisk	-0.38184 (0.32519)	0.019468 (0.0159042)	0.0206435 (0.016227)	0.0049986 (0.039466)	0.0219464 (0.015177)	0.017718 (0.016990)
Devel	-0.22558 (0.27952)	-0.0168784 (0.033961)	-0.016135 (0.03388)	0.050883 (0.077570)	-0.0166715 (0.030232)	-0.017778 (0.035999)
lambda	na	na	na	na	0.0576486 (0.065882)	na

*LIV and $Bayes_2$ plagued with convergence issues; both remain works in progress

Parameter estimates for outcome equations

D = 0

Regressor	LIV*	$Bayes_1$	$Bayes_2$ .79	$Bayes_2$ .21	H2SCF	OLS
Intercept	na	-1.848865 (0.808426)	-2.149203 (0.925848)	0.30967 (0.430852)	-4.7993693 (3.366629)	-1.826023 (0.37904)
Prestige 1	na	-0.187058 (0.125690)	-0.603723 (0.182573)	0.0424368 (0.053957)	0.0519449 (0.276472)	-0.18902 (0.06282)
Prestige 2	na	-0.192381 (0.170706)	-4.88652 (0.241644)	0.130233 (0.080528)	0.0631206 (0.345934)	-0.23741 (0.08824)
LnAssets	-.007300 (0.04391)	0.012546 (0.021016)	-0.003911 (0.028636)	-0.039873 (0.012874)	-0.0078691 (0.033752)	0.01307 (0.01037)
LnOffVal	-0.38877 (0.30730)	0.118618 (0.067253)	0.132283 (0.066665)	0.033753 (0.032976)	0.3623773 (0.277965)	0.11481 (0.02778)
IndRisk	0.439375 (0.34444)	-0.190472 (0.103841)	-2.07682 (0.169247)	0.084637 (0.049924)	-0.0951526 (0.162607)	-0.20582 (0.05207)
Devel	0.27622 (0.37619)	-0.167953 (0.111247)	0.094308 (0.087937)	2.085604 (0.537663)	-0.1246316 (0.154355)	-0.18406 (0.05678)
lambda	na	na	na	na	0.5613458 (0.596482)	na

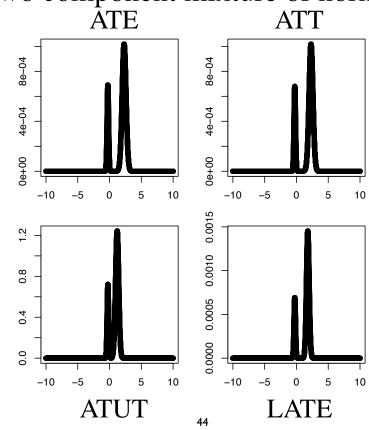
*LIV and $Bayes_2$ plagued with convergence issues; both remain works in progress

Estimated treatment effects

	LIV	Bayes$_1$ {Pr(TE>0)}	Bayes$_2$ {Pr(TE>0)} [modes]	H2SCF	OLS
ATE	na	-0.090899 {0.430355}	1.753479 {0.7925126} [-0.265, 2.29]	-1.091951	-0.1797867
ATT	na	-0.096009 {0.427659}	1.75384 {0.7929212} [-0.264, 2.29]	-1.17614	na
ATUT	na	-0.036403 {0.458698}	0.8873612 {0.7916477} [-0.263, 1.20]	-0.1650830	na
LATE [0.80,0.85] [0.94,0.95]*	na	-0.063626 { 0.428764}	1.357273 {0.792181} [-0.265, 1.785]	-0.5247407	na

*Bayes$_2$ interval for LATE

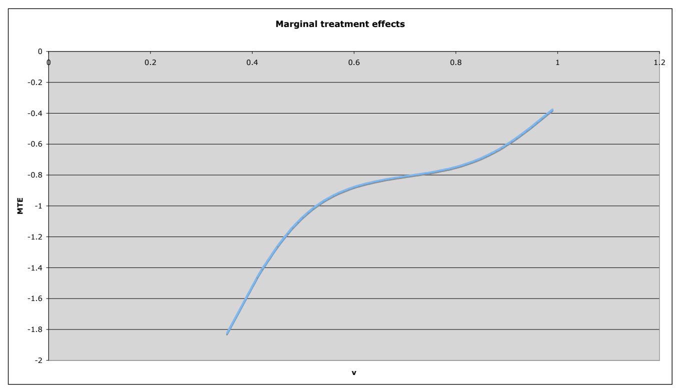
Predictive density plots for treatment effects based on two component mixture of normals



Estimated treatment effects
excluding PREST1 & PREST2

	LIV	Bayes$_1$ {Pr(TE>0)}	Bayes$_2$ {Pr(TE>0)}	H2SCF	OLS
ATE	-0.72438 (0.65664)	-0.59984 {0.08191}	-0.096605 {0.357392}	-0.9114187	-0.24632
ATT	-0.77383 (0.73560)	-0.64712 {0.051372}	-0.10371 {0.3573354}	-0.9774227	na
ATUT	-0.46402 (0.39799)	-0.05219 {0.43413}	na	-0.1847533	na
LATE [0.80,0.85]	-68927 (0.55006)	-0.30092 {0.168199} [0.81, 0.86]	na	-0.4667252	na

Nonlinearity in MTE



Conclusions

Accounting is heavily populated with choices

Accounting choices are (almost surely) endogenous

Progress in understanding the role of accounting in the world requires care and creativity

– richer, more elegant *theory*

– improved *data* (attention to omitted correlated variables, heterogeneity, and instruments)

– more extensive *diagnostic checking/discovery of the DGP*

No algorithm exists or is likely to in the future.