Contents

9
Bayesian causal effects strategies

In the chapter we revisit causal and treatment effects but instead of appealing to classical strategies we explore some Bayesian strategies. For instance, Bayesian data augmentation might replace the classical control or replacement function.

9.1 Treatment effects and counterfactuals

Suppose we observe treatment or no treatment and the associated outcome, $Y = DY_1 + (1 - D)\,Y_0$, where

$$\begin{aligned} Y_1 &= \beta_1 + V_1 \\ Y_0 &= \beta_0 + V_0 \end{aligned}$$

and a representative sample is

Y	D	Y_1	Y_0	V_1	V_0
15	1	15	9	3	-3
14	1	14	10	2	-2
13	1	13	11	1	-1
13	0	11	13	-1	1
14	0	10	14	-2	2
15	0	9	15	-3	3

Further, we have the following instruments at our disposal $Z = \begin{bmatrix} Z_1 & Z_2 & Z_3 & Z_4 \end{bmatrix}$ where their representative values are

Z_1	Z_2	Z_3	Z_4
5	4	3	1
-6	-5	-4	-2
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	0

and we perceive latent utility, EU, to be related to choice via the instruments.

$$EU = Z\theta + V_D$$

and observed choice is

$$D = \begin{cases} 1 & EU > 0 \\ 0 & \text{otherwise} \end{cases}$$

This is the exact setup we discussed earlier in the projections analysis.

9.2 Posterior distribution

Define the complete or augmented data as

$$r_i = \begin{bmatrix} D_i^* & D_i Y_i + (1 - D_i) Y_i^{miss} & D_i Y_i^{miss} + (1 - D_i) Y_i \end{bmatrix}^T$$

Also, let

$$H_i = \begin{bmatrix} Z_i & 0 & 0 \\ 0 & X_i & 0 \\ 0 & 0 & X_i \end{bmatrix}$$

and

$$\beta = \begin{bmatrix} \theta \\ \beta_1 \\ \beta_0 \end{bmatrix}$$

where X is a matrix of outcome regressors, in the current example it is simply ι, a vector of ones, as there are no outcome covariates. Hence, a compact model is

$$r_i = H_i \beta + \varepsilon_i$$

where $\varepsilon_i = \begin{bmatrix} V_{Di} \\ V_{1i} \\ V_{0i} \end{bmatrix}$ and $\Sigma = Var[\varepsilon_i] = \begin{bmatrix} 1 & \sigma_{D1} & \sigma_{D0} \\ \sigma_{D1} & \sigma_1^2 & \sigma_{10} \\ \sigma_{D0} & \sigma_{10} & \sigma_0^2 \end{bmatrix}$.

9.2.1 *Likelihood function*

As usual the posterior distribution is proportional to the likelihood function times the prior distribution. The likelihood function is

$$r_i \sim N\left(H_i\beta, \Sigma\right)$$

Or,

$$\ell\left(\beta, \Sigma \mid r_i, D_i, X_i, Z_i\right) \propto \left|\Sigma\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(r_i - H_i\beta\right)^T \Sigma^{-1}\left(r_i - H_i\beta\right)\right]$$

9.2.2 *Prior distribution*

Frequently, relatively diffuse priors are chosen such that the data dominates the posterior distribution. Li, Poirier, and Tobias' prior distribution for β is $p\left(\beta\right) \sim N\left(\beta_0, V_\beta\right)$ where $\beta_0 = 0, V_\beta = 4I$ and their independent prior for Σ^{-1} is $p\left(\Sigma^{-1}\right) \sim Wishart\left(\rho, \rho R\right)$ or for Σ is $p\left(\Sigma\right) \sim InverseWishart\left(\rho, \left(\rho R\right)^{-1}\right)$where $\rho = 12$ and R is a diagonal matrix with elements $\left\{\frac{1}{12}, \frac{1}{4}, \frac{1}{4}\right\}$. Hence, the joint conjugate prior is normal-inverse Wishart.

$$\begin{aligned} p\left(\beta, \Sigma\right) &= p\left(\beta\right) p\left(\Sigma\right) \\ &\propto \left|V_\beta\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\beta - \beta_0\right)^T V_\beta^{-1}\left(\beta - \beta_0\right)\right] \\ &\quad \times \left|\Sigma\right|^{-\frac{\rho+4}{2}} \exp\left[-\frac{1}{2}Tr\left(\rho R\Sigma^{-1}\right)\right] \end{aligned}$$

where $Tr\left(\cdot\right)$ is the trace of the matrix.

9.2.3 *Posterior distribution*

Now, the posterior distribution (or posterior kernel) is

$$p\left(\beta, \Sigma, Y_i^{miss}, D_i^* \mid Y_i, D_i, X_i, Z_i\right) \propto p\left(\beta, \Sigma\right) \ell\left(\beta, \Sigma \mid r_i, D_i, X_i, Z_i\right)$$

9.3 Gibbs sampler for treatment effects

As is frequently the case, it's much easier to simulate from the recognizable conditional posterior distributions via a Gibbs sampler than simulate from the unrecognizable joint posterior distribution. There are three sources of missing data: latent utility, EU, counterfactuals for individuals who choose treatment, $\left(Y_{0i} \mid D_i = 1\right)$, and counterfactuals for individuals who choose no treatment, $(Y_{1i} \mid D_i = 0)$. Bayesian data augmentation effectively models these missing data processes (as in, for example, Albert and Chib's *McMC* probit) by drawing in sequence from the conditional posterior distributions — a Gibbs sampler.

9.3.1 Full conditional posterior distributions

First block

Let Γ_{-x} denote all parameters other than x. The full conditional posteriors for the augmented outcome data are

$$\begin{aligned} p\left(Y_i^{miss} \mid \Gamma_{-Y_i^{miss}}, Y_i, D_i, X_i, Z_i\right) &\propto \frac{p\left(\beta, \Sigma, Y_i^{miss}, D_i^* \mid Y_i, D_i, X_i, Z_i\right)}{p\left(\beta, \Sigma\right) p\left(Y_i, D_i^* \mid D_i, X_i, Z_i\right)} \\ &\propto \frac{\ell\left(\beta, \Sigma \mid r_i, D_i, X_i, Z_i\right)}{p\left(Y_i, D_i^* \mid D_i, X_i, Z_i\right)} \end{aligned}$$

Hence,

$$Y_i^{miss} \mid \Gamma_{-Y_i^{miss}}, Y_i, D_i, X_i, Z_i \sim N\left(Y_i^{miss} \mid Y_i, D_i^*, D_i, X_i, Z_i; \beta, \Sigma\right)$$

In other words, the posterior for the missing data is normal conditional on observed outcome, Y_i, and latent expected utility, D_i^*. Standard multivariate normal theory (see the appendix) provides the means and variances conditional on the draw for latent utility and the other outcome.

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

where $Data$ refers to (Y_i, D_i, X_i, Z_i)

$$\mu_{1i} = X_i \beta_1 + \frac{\sigma_0^2 \sigma_{D1} - \sigma_{10} \sigma_{D0}}{\sigma_0^2 - \sigma_{D0}^2} \left(D_i^* - Z_i \theta\right) + \frac{\sigma_{10} - \sigma_{D1} \sigma_{D0}}{\sigma_0^2 - \sigma_{D0}^2} \left(Y_i - X_i \beta_0\right)$$

$$\mu_{0i} = X_i \beta_0 + \frac{\sigma_1^2 \sigma_{D0} - \sigma_{10} \sigma_{D1}}{\sigma_1^2 - \sigma_{D1}^2} \left(D_i^* - Z_i \theta\right) + \frac{\sigma_{10} - \sigma_{D1} \sigma_{D0}}{\sigma_1^2 - \sigma_{D1}^2} \left(Y_i - X_i \beta_1\right)$$

$$\omega_{1i} = \sigma_1^2 - \frac{\sigma_{D1}^2 \sigma_0^2 - 2 \sigma_{10} \sigma_{D1} \sigma_{D0} + \sigma_{10}^2}{\sigma_0^2 - \sigma_{D0}^2}$$

$$\omega_{0i} = \sigma_0^2 - \frac{\sigma_{D0}^2 \sigma_1^2 - 2 \sigma_{10} \sigma_{D1} \sigma_{D0} + \sigma_{10}^2}{\sigma_1^2 - \sigma_{D1}^2}$$

Similarly, the conditional posterior for latent expected utility is

$$\begin{aligned} p\left(D_i^* \mid \Gamma_{-D_i^*}, Y_i, D_i, X_i, Z_i\right) &\propto \frac{p\left(\beta, \Sigma, Y_i^{miss}, D_i^* \mid Y_i, D_i, X_i, Z_i\right)}{p\left(\beta, \Sigma\right) p\left(Y_i, Y_i^{miss} \mid D_i, X_i, Z_i\right)} \\ &\propto \frac{\ell\left(\beta, \Sigma \mid r_i, D_i, X_i, Z_i\right)}{p\left(Y_i, Y_i^{miss} \mid D_i, X_i, Z_i\right)} \end{aligned}$$

Hence,

$$D_i^* \mid \Gamma_{-D_i^*}, Y_i, D_i, X_i, Z_i \sim N\left(D_i^* \mid Y_i, Y_i^{miss}, D_i, X_i, Z_i; \beta, \Sigma\right)$$

In other words, the posterior for latent expected utility is truncated normal conditioned on observed and missing outcomes.

$$D_i^* \mid \Gamma_{-D_i^*}, Data \sim \begin{array}{ll} TN_{(0,\infty)}\left(\mu_{D_i}\omega_D\right) & if\ D_i = 1 \\ TN_{(-\infty,0)}\left(\mu_{D_i}\omega_D\right) & if\ D_i = 0 \end{array}$$

where $TN(\cdot)$ refers to the truncated normal distribution with support indicated via the subscript and the arguments are parameters of the untruncated distribution. Applying multivariate normal theory for $(D_i^* \mid Y_i)$ we have

$$\begin{aligned} \mu_{D_i} &= Z_i\theta + \left(D_iY_i + (1 - D_i)\,Y_i^{miss} - X_i\beta_1\right)\frac{\sigma_0^2\sigma_{D1} - \sigma_{10}\sigma_{D0}}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2} \\ &\quad + \left(D_iY_i^{miss} + (1 - D_i)\,Y_i - X_i\beta_0\right)\frac{\sigma_1^2\sigma_{D0} - \sigma_{10}\sigma_{D1}}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2} \end{aligned}$$

$$\omega_D = 1 - \frac{\sigma_{D1}^2\sigma_0^2 - 2\sigma_{10}\sigma_{D1}\sigma_{D0} + \sigma_{D0}^2\sigma_1^2}{\sigma_1^2\sigma_0^2 - \sigma_{10}^2}$$

Second block

With prior distribution $p(\beta) \sim N(\beta_0, V_\beta)$, the conditional posterior distribution for the parameters is

$$\begin{aligned} p\left(\beta \mid \Gamma_{-\beta}, Y_i, D_i, X_i, Z_i\right) &\propto \frac{p\left(\beta, \Sigma, Y_i^{miss}, D_i^* \mid Y_i, D_i, X_i, Z_i\right)}{p(\Sigma)\,p\left(Y_i, Y_i^{miss}, D_i^* \mid D_i, X_i, Z_i\right)} \\ &\propto \frac{p(\beta)\,\ell\left(\beta, \Sigma \mid r_i, D_i, X_i, Z_i\right)}{p\left(Y_i, Y_i^{miss}, D_i^* \mid D_i, X_i, Z_i\right)} \end{aligned}$$

In other words, the posterior for the parameters is normal conditioned on observed and missing outcomes, latent expected utility, and variance Σ.

$$\beta \mid \Gamma_{-\beta}, Data \sim N\left(\mu_\beta, \omega_\beta\right)$$

where by the *SUR* (seemingly-unrelated regression) generalization of Bayesian regression (see the appendix)

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

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

With prior $p(\Sigma) \sim Wishart(\rho, \rho R)$, the conditional distribution for the trivariate variance-covariance matrix is

$$\begin{aligned} p(\Sigma \mid \Gamma_{-\Sigma}, Y_i, D_i, X_i, Z_i) &\propto \frac{p\left(\beta, \Sigma, Y_i^{miss}, D_i^* \mid Y_i, D_i, X_i, Z_i\right)}{p(\beta)\, p\left(Y_i, Y_i^{miss}, D_i^* \mid D_i, X_i, Z_i\right)} \\ &\propto \frac{p(\Sigma)\, \ell(\beta, \Sigma \mid r_i, D_i, X_i, Z_i)}{p\left(Y_i, Y_i^{miss}, D_i^* \mid D_i, X_i, Z_i\right)} \end{aligned}$$

Hence,

$$\Sigma \mid \Gamma_{-\Sigma}, Y_i, D_i, X_i, Z_i \sim Wishart\left(\Sigma \mid Y_i, Y_i^{miss}, D_i^*, D_i, X_i, Z_i, \Sigma; \beta_0, V_\beta\right)$$

In other words, the posterior for the parameters is inverse-Wishart conditioned on observed and missing outcomes, latent expected utility, and parameters β.

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

where

$$G \sim Wishart(n + \rho, S + \rho R)$$

and $S = \sum_{i=1}^{n} (r_i - H_i\beta)(r_i - H_i\beta)^T$.[1]

As usual, starting values for the Gibbs sampler are varied to test convergence of the posterior distributions (adequate coverage of the sample space). Stationary convergence plots and quickly dampening autocorrelation plots support the notion of representative posterior draws.

9.3.2 Nobile's algorithm

Recall σ_D^2 is normalized to one. This creates a slight complication as the conditional posterior is no longer inverse-Wishart. Nobile [2000] provides a convenient algorithm for random Wishart (multivariate χ^2) draws with a restricted element. The algorithm applied to the current setting results in the following steps:

1. Exchange rows and columns one and three in $S + \rho R$, call this matrix V.
2. Find L such that $V = \left(L^{-1}\right)^T L^{-1}$.

[1] Technically, σ_{10} is unidentified (i.e., even with unlimited data we cannot "observe" the parameter). However, we can employ restrictions derived through the positive-definiteness (see the appendix) of the variance-covariance matrix, Σ, to impose bounds on the parameter, σ_{10}. If treatment effects are overly sensitive this strategy will prove ineffective; otherwise, it allows us to proceed from observables to treatment effects via augmentation of unobservables (the counterfactuals as well as latent utility).

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. $a_{33} = \frac{1}{l_{33}}$ where l_{33} is the third row-column element of L.
   c. a_{ij} equal to $N(0, 1)$ random variates, $i > j$.

4. Set $V' = \left(L^{-1}\right)^T \left(A^{-1}\right)^T A^{-1} L^{-1}$.

5. Exchange rows and columns one and three in V' and denote this draw Σ.

9.4 Marginal and average treatment effects

The marginal treatment effect is the impact of treatment for individuals who are indifferent between treatment and no treatment. We can employ Bayesian data augmentation-based estimation of marginal treatment effects (MTE) as data augmentation generates repeated draws for unobservables, V_{Dj}, $(Y_{1j} \mid D_j = 0)$, and $(Y_{0j} \mid D_j = 1)$. Now, exploit these repeated samples to describe the distribution for $MTE(u_D)$ where V_D is transformed to uniform $(0, 1)$, $u_D = p_v$. For each draw, $V_D = v$, we determine the cumulative probability, $u_D = \Phi(v)$,[2] and calculate $MTE(u_D) = E[Y_1 - Y_0 \mid u_D]$. If $MTE(u_D)$ is constant for all u_D, then all treatment effects are alike.

MTE can be connected to standard population-level treatment effects, ATE, ATT, and $ATUT$, via non-negative weights whose sum is one (assuming full support)

$$\begin{aligned}
w_{ATE}(u_D) &= \frac{\sum_{j=1}^{n} I(u_D)}{n} \\
w_{ATT}(u_D) &= \frac{\sum_{j=1}^{n} I(u_D) D_j}{\sum_{j=1}^{n} D_j} \\
w_{ATUT}(u_D) &= \frac{\sum_{j=1}^{n} I(u_D)(1 - D_j)}{\sum_{j=1}^{n} (1 - D_j)}
\end{aligned}$$

where probabilities p_k refer to bins from 0 to 1 by increments of 0.01 for indicator variable

$$\begin{array}{ll}
I(u_D) = 1 & u_D = p_k \\
I(u_D) = 0 & u_D \neq p_k
\end{array}$$

[2] $\Phi(\cdot)$ is a cumulative probability distribution function.

Hence, MTE-estimated average treatment effects are

$$\begin{aligned}
estATE\left(MTE\right) &= \sum_{i=1}^{n} w_{ATE}\left(u_D\right) MTE\left(u_D\right) \\
estATT\left(MTE\right) &= \sum_{i=1}^{n} w_{ATT}\left(u_D\right) MTE\left(u_D\right) \\
estATUT\left(MTE\right) &= \sum_{i=1}^{n} w_{ATUT}\left(u_D\right) MTE\left(u_D\right)
\end{aligned}$$

Next, we apply these data augmentation ideas to the causal effects example and estimate the average treatment effect on the treated (ATT), the average treatment effect on the untreated $(ATUT)$, and the average treatment effect (ATE).

9.5 Return to the treatment effect example

Initially, we employ Bayesian data augmentation via a Gibbs sampler on the treatment effect problem outlined above. Recall this example was employed in the projections notes to illustrate where the inverse-Mills ratios control functions strategy based on the full complement of instruments[3] was exceptionally effective.

The representative sample is

Y	D	Y_1	Y_0	Z_1	Z_2	Z_3	Z_4
15	1	15	9	5	4	3	1
14	1	14	10	-6	-5	-4	-2
13	1	13	11	0	0	0	1
13	0	11	13	0	0	1	0
14	0	10	14	0	1	0	0
15	0	9	15	1	0	0	0

which is repeated 200 times to create a sample of $n = 1,200$ observations. The Gibbs sampler employs $15,000$ draws from the conditional posteriors. The first $5,000$ draws are discarded as burn-in, then sample statistics are

[3] Typically, we're fortunate to identify any instruments. In the example, the instruments form a basis for the nullspace to the outcomes, Y_1 and Y_0. In this (linear or Gaussian) sense, we've exhausted the potential set of instruments.

based on the remaining $10,000$ draws.

statistic	β_1	β_0	θ_1	θ_2	θ_3	θ_4
mean	13.76	13.76	0.810	−0.391	−1.647	1.649
median	13.76	13.76	0.809	−0.391	−1.645	1.650
standard dev	0.026	0.028	0.051	0.054	0.080	0.080
quantiles:						
minimum	13.67	13.64	0.617	−0.585	−1.943	1.362
0.01	13.70	13.69	0.695	−0.521	−1.837	1.461
0.025	13.71	13.70	0.713	−0.500	−1.807	1.493
0.05	13.72	13.71	0.727	−0.481	−1.781	1.518
0.10	13.73	13.71	0.746	−0.461	−1.751	1.547
0.25	13.74	13.74	0.776	−0.428	−1.699	1.595
0.75	13.78	13.78	0.844	−0.356	−1.593	1.704
0.90	13.79	13.80	0.873	−0.325	−1.547	1.751
0.95	13.80	13.80	0.893	−0.306	−1.519	1.778
0.975	13.81	13.81	0.910	−0.289	−1.497	1.806
0.99	13.82	13.82	0.931	−0.269	−1.467	1.836
maximum	13.84	13.86	1.006	−0.185	−1.335	1.971

Sample statistics for the parameters of the data augmented Gibbs sampler applied to the treatment effect example

The results demonstrate selection bias as the means are biased upward from 12. This does not bode well for effective estimation of marginal or average treatment effects. Sample statistics for average treatment effects as well as correlations, $\rho_{D,1}$, $\rho_{D,0}$, and $\rho_{1,0}$ are tabulated below.

statistic	ATE	ATT	$ATUT$	$\rho_{D,1}$	$\rho_{D,0}$	$\rho_{1,0}$
mean	−0.000	0.481	−0.482	0.904	−0.904	−0.852
median	−0.000	0.480	−0.481	0.904	−0.904	−0.852
standard dev	0.017	0.041	0.041	0.009	0.009	0.015
quantiles:						
minimum	−0.068	0.331	−0.649	0.865	−0.933	−0.899
0.01	−0.039	0.388	−0.580	0.880	−0.923	−0.884
0.025	−0.033	0.403	−0.564	0.884	−0.920	−0.879
0.05	−0.028	0.415	−0.549	0.888	−0.918	−0.875
0.10	−0.022	0.428	−0.534	0.892	−0.915	−0.871
0.25	−0.012	0.452	−0.509	0.898	−0.910	−0.862
0.75	0.011	0.510	−0.453	0.910	−0.898	−0.842
0.90	0.022	0.535	−0.429	0.915	−0.892	−0.832
0.95	0.028	0.551	−0.416	0.917	−0.888	−0.826
0.975	0.034	0.562	−0.405	0.920	−0.884	−0.821
0.99	0.040	0.576	−0.393	0.923	−0.880	−0.814
maximum	0.068	0.649	−0.350	0.932	−0.861	−0.787

Sample statistics for average treatment effects and error correlations of the data augmented Gibbs sampler applied to the treatment effect example

Average treatment effects estimated from weighted averages of MTE are similar:

$$\begin{aligned} estATE\left(MTE\right) &= -0.000 \\ estATT\left(MTE\right) &= 0.464 \\ estATUT\left(MTE\right) &= -0.464 \end{aligned}$$

The average treatment effects on the treated and untreated suggest heterogeneity but are grossly understated compared to the DGP averages of 4 and -4. Next, we revisit the problem and attempt to consider what is left out of our model specification.

9.6 Instrumental variable restrictions

Consistency demands that we fully consider what we know. In the foregoing analysis, we have not effectively employed this principle. Data augmentation of the counterfactuals involves another condition. That is, outcomes are independent of the instruments (otherwise, they are not instruments), $DY + (1-D)\,Y^{draw}$ and $DY^{draw} + (1-D)\,Y$ are independent of Z. We can impose orthogonality on the draws of the counterfactuals such that the "sample" satisfies this population condition.[4] We'll refer to this as the IV data augmented Gibbs sampler treatment effect analysis.

To implement this we add the following steps to the above Gibbs sampler. Minimize the distance of Y^{draw} from Y^{miss} such that $Y_1^* = DY + (1-D)\,Y^{draw}$ and $Y_0^* = DY^{draw} + (1-D)\,Y$ are orthogonal to the instruments, Z.

$$\begin{aligned} &\min_{Y^{draw}} \left(Y^{draw} - Y^{miss}\right)^T \left(Y^{draw} - Y^{miss}\right) \\ &s.t. \quad Z^T \left[\begin{array}{cc} DY + (1-D)\,Y^{draw} & DY^{draw} + (1-D)\,Y \end{array}\right] = 0 \end{aligned}$$

where the constraint is $p \times 2$ zeroes and p is the number of columns in Z (the number of instruments). Hence, the IV $McMC$ outcome draws are

$$Y_1^* = DY + (1-D)\,Y^{draw}$$

and

$$Y_0^* = DY^{draw} + (1-D)\,Y$$

[4] Whenever observed data fails to provide broad coverage of the sample space instrumentation alone is likely to be ineffective. In this case, we're hoping to exploit the instruments via the algorithm to identify counterfactuals (unobservable data) and model parameters. With sparse coverage we can assist the algorithm if we have a rich set of instruments available.

9.7 Return to the example once more

With the IV data augmented Gibbs sampler in hand we return to the representative sample

Y	D	Y_1	Y_0	Z_1	Z_2	Z_3	Z_4
15	1	15	9	5	4	3	1
14	1	14	10	−6	−5	−4	−2
13	1	13	11	0	0	0	1
13	0	11	13	0	0	1	0
14	0	10	14	0	1	0	0
15	0	9	15	1	0	0	0

and repeat 20 times to create a sample of $n = 120$ observations. The IV Gibbs sampler employs 15,000 draws from the conditional posteriors. The first 5,000 draws are discarded as burn-in, then sample statistics are based on the remaining 10,000 draws.

statistic	β_1	β_0	θ_1	θ_2	θ_3	θ_4
mean	12.01	11.99	0.413	−0.167	−0.896	0.878
median	12.01	11.99	0.420	−0.148	−0.866	0.852
standard dev	0.160	0.160	0.227	0.274	0.370	0.359
quantiles:						
minimum	11.35	11.37	−0.558	−1.325	−2.665	−0.202
0.01	11.64	11.62	−0.149	−0.889	−1.888	0.170
0.025	11.69	11.68	−0.058	−0.764	−1.696	0.254
0.05	11.74	11.73	0.028	−0.648	−1.550	0.336
0.10	11.80	11.80	0.117	−0.530	−1.381	0.435
0.25	11.90	11.89	0.267	−0.334	−1.124	0.617
0.75	12.11	12.10	0.566	0.023	−0.637	1.113
0.90	12.21	12.20	0.695	0.168	−0.451	1.367
0.95	12.27	12.25	0.774	0.249	−0.348	1.509
0.975	12.32	12.30	0.840	0.312	−0.256	1.630
0.99	12.38	12.36	0.923	0.389	−0.170	1.771
maximum	12.63	12.64	1.192	0.685	0.257	2.401

Sample statistics for the parameters of the IV data augmented Gibbs sampler applied to the treatment effect example

Not surprisingly, the results demonstrate no selection bias and effectively estimate marginal and average treatment effects. Sample statistics for average treatment effects as well as correlations, $\rho_{D,1}$, $\rho_{D,0}$, and $\rho_{1,0}$ are tabulated below.

statistic	ATE	ATT	$ATUT$	$\rho_{D,1}$	$\rho_{D,0}$	$\rho_{1,0}$
mean	0.000	4.000	-4.000	0.813	-0.812	-0.976
median	0.000	4.000	-4.000	0.815	-0.815	-0.976
standard dev	0.000	0.000	0.000	0.031	0.032	0.004
quantiles:						
minimum	0.000	4.000	-4.000	0.650	-0.910	-0.987
0.01	0.000	4.000	-4.000	0.728	-0.874	-0.984
0.025	0.000	4.000	-4.000	0.743	-0.866	-0.983
0.05	0.000	4.000	-4.000	0.756	-0.859	-0.982
0.10	0.000	4.000	-4.000	0.772	-0.851	-0.981
0.25	0.000	4.000	-4.000	0.794	-0.835	-0.979
0.75	0.000	4.000	-4.000	0.835	-0.794	-0.973
0.90	0.000	4.000	-4.000	0.850	-0.771	-0.970
0.95	0.000	4.000	-4.000	0.859	-0.755	-0.968
0.975	0.000	4.000	-4.000	0.866	-0.742	-0.967
0.99	0.000	4.000	-4.000	0.874	-0.726	-0.965
maximum	0.000	4.000	-4.000	0.904	-0.640	-0.952

Sample statistics for average treatment effects and error correlations of the IV data augmented Gibbs sampler applied to the treatment effect example

Weighted MTE estimates of average treatment effects are similar.

$estATE\,(MTE)$	$estATT\,(MTE)$	$estATUT\,(MTE)$
0.000	3.792	-3.792

Next, we report some more interesting experiments. Instead, of having the full set of instruments available, suppose we have only three, Z_1, Z_2, and $Z_3 + Z_4$, or two, $Z_1 + Z_2$ and $Z_3 + Z_4$, or one, $Z_1 + Z_2 + Z_3 + Z_4$. We repeat the above for each set of instruments and compare the results with classical control function analysis based on Heckman's inverse Mills strategy introduced in the projections notes.

9.7.1 Three instruments

Suppose we have only three instruments, Z_1, Z_2, and $Z_3 + Z_4$. *IV* data augmented Gibbs sampler results are tabulated below.[5]

statistic	β_1	β_0	θ_1	θ_2	θ_3
mean	12.00	12.00	0.242	−0.358	−0.001
median	12.00	12.00	0.243	−0.342	−0.001
standard dev	0.164	0.165	0.222	0.278	0.132
quantiles:					
minimum	11.32	11.36	−0.658	−1.451	−0.495
0.01	11.62	11.61	−0.263	−1.080	−0.306
0.025	11.68	11.68	−0.189	−0.950	−0.258
0.05	11.73	11.73	−0.120	−0.844	−0.216
0.10	11.79	11.79	−0.041	−0.723	−0.170
0.25	11.89	11.89	0.094	−0.532	−0.091
0.75	12.11	12.11	0.394	−0.168	0.090
0.90	12.21	12.21	0.526	−0.021	0.171
0.95	12.27	12.27	0.604	0.071	0.217
0.975	12.32	12.32	0.670	0.155	0.254
0.99	12.38	12.39	0.753	0.245	0.302
maximum	12.58	12.57	1.067	0.564	0.568

Sample statistics for the parameters of the *IV* data augmented Gibbs sampler with three instruments applied to the treatment effect example

These results differ very little from those based on the full set of four instruments. There is no selection bias and marginal and average treatment effects are effectively estimated. Sample statistics for average treatment

[5] Inclusion of an intercept in the selection equation with three, two, and one instruments makes no qualitative difference in the average treatment effect analysis. These results are not reported.

effects as well as correlations, $\rho_{D,1}$, $\rho_{D,0}$, and $\rho_{1,0}$ are tabulated below.

statistic	ATE	ATT	$ATUT$	$\rho_{D,1}$	$\rho_{D,0}$	$\rho_{1,0}$
mean	0.000	4.000	−4.000	0.799	−0.800	−0.884
median	0.000	4.000	−4.000	0.802	−0.814	−0.888
standard dev	0.000	0.000	0.000	0.036	0.037	0.029
quantiles:						
minimum	0.000	4.000	−4.000	0.605	−0.899	−0.956
0.01	0.000	4.000	−4.000	0.702	−0.870	−0.936
0.025	0.000	4.000	−4.000	0.719	−0.861	−0.930
0.05	0.000	4.000	−4.000	0.734	−0.853	−0.924
0.10	0.000	4.000	−4.000	0.751	−0.844	−0.918
0.25	0.000	4.000	−4.000	0.777	−0.826	−0.905
0.75	0.000	4.000	−4.000	0.825	−0.778	−0.867
0.90	0.000	4.000	−4.000	0.842	−0.751	−0.846
0.95	0.000	4.000	−4.000	0.852	−0.734	−0.833
0.975	0.000	4.000	−4.000	0.860	−0.720	−0.821
0.99	0.000	4.000	−4.000	0.869	−0.699	−0.803
maximum	0.000	4.000	−4.000	0.894	−0.554	−0.703

Sample statistics for average treatment effects and error correlations of the IV data augmented Gibbs sampler with three instruments applied to the treatment effect example

Weighted MTE estimates of average treatment effects are similar.

$estATE\,(MTE)$	$estATT\,(MTE)$	$estATUT\,(MTE)$
−0.000	3.940	−3.940

Classical results based on Heckman's inverse Mills control function strategy with three instruments are reported below for comparison. The selection equation estimated via probit is

$$\Pr(D \mid Z) = \Phi(0.198Z_1 - 0.297Z_2 + 0.000(Z_3 + Z_4)) \quad PseudoR^2 = 0.019$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function. The estimated outcome equations are

$$E[Y \mid X] = 11.890(1 - D) + 11.890D - 2.700(1 - D)\lambda_0 + 2.700D\lambda_1$$

and estimated average treatment effects are

$estATE$	$estATT$	$estATUT$
0.000	4.220	−4.220

In spite of the weak explanatory of the selection model, control functions produce reasonable estimates of average treatment effects. Next, we consider two instruments.

9.7.2 Two instruments

Suppose we have only two instruments, $Z_1 + Z_2$, and $Z_3 + Z_4$. *IV* data augmented Gibbs sampler results are tabulated below.

statistic	β_1	β_0	θ_1	θ_2
mean	12.08	13.27	-0.034	0.008
median	12.07	13.27	-0.034	0.009
standard dev	0.168	0.243	0.065	0.128
quantiles:				
minimum	11.47	12.41	-0.328	-0.579
0.01	11.69	12.70	-0.185	-0.287
0.025	11.75	12.79	-0.162	-0.244
0.05	11.80	12.87	-0.141	-0.207
0.10	11.86	12.96	-0.118	-0.159
0.25	11.96	13.11	-0.077	-0.077
0.75	12.18	13.42	0.009	0.095
0.90	12.29	13.58	0.048	0.171
0.95	12.35	13.67	0.073	0.219
0.975	12.41	13.75	0.092	0.260
0.99	12.46	13.84	0.115	0.308
maximum	12.64	14.26	0.260	0.635

Sample statistics for the parameters of the *IV* data augmented Gibbs sampler with two instruments applied to the treatment effect example

Selection bias emerges as β_0 diverges from 12. This suggests marginal and average treatment effects are likely to be confounded. Sample statistics for average treatment effects as well as correlations, $\rho_{D,1}$, $\rho_{D,0}$, and $\rho_{1,0}$ are tabulated below.

statistic	ATE	ATT	$ATUT$	$\rho_{D,1}$	$\rho_{D,0}$	$\rho_{1,0}$
mean	-1.293	1.413	-4.000	0.802	-0.516	-0.634
median	-1.297	1.406	-4.000	0.806	-0.532	-0.648
standard dev	0.219	0.438	0.000	0.037	0.136	0.115
quantiles:						
minimum	-2.105	-0.211	-4.000	0.601	-0.813	-0.890
0.01	-1.806	0.389	-4.000	0.695	-0.757	-0.834
0.025	-1.738	0.525	-4.000	0.719	-0.732	-0.813
0.05	-1.665	0.670	-4.000	0.735	-0.706	-0.795
0.10	-1.572	0.855	-4.000	0.754	-0.675	-0.768
0.25	-1.435	1.130	-4.000	0.779	-0.613	-0.716
0.75	-1.147	1.705	-4.000	0.828	-0.438	-0.569
0.90	-1.005	1.989	-4.000	0.846	-0.340	-0.479
0.95	-0.930	2.141	-4.000	0.856	-0.262	-0.417
0.975	-0.861	2.277	-4.000	0.864	-0.195	-0.365
0.99	-0.795	2.409	-4.000	0.874	-0.124	-0.301
maximum	-0.625	2.750	-4.000	0.902	0.150	-0.055

Sample statistics for average treatment effects and error correlations of the IV data augmented Gibbs sampler with two instruments applied to the treatment effect example

Weighted MTE estimates of average treatment effects are similar.

$estATE\,(MTE)$	$estATT\,(MTE)$	$estATUT\,(MTE)$
-1.293	1.372	-3.959

$ATUT$ is effectively estimated but the other average treatment effects are biased.

Classical results based on Heckman's inverse Mills control function strategy with two instruments are reported below for comparison. The selection equation estimated via probit is

$$\Pr\left(D \mid Z\right) = \Phi\left(-0.023\left(Z_1 + Z_2\right) + 0.004\left(Z_3 + Z_4\right)\right) \quad PseudoR^2 = 0.010$$

The estimated outcome equations are

$$E\left[Y \mid X\right] = 109.38\left(1 - D\right) + 11.683D + 121.14\left(1 - D\right)\lambda_0 + 2.926D\lambda_1$$

and estimated average treatment effects are

$estATE$	$estATT$	$estATUT$
-97.69	-191.31	-4.621

While the Bayesian estimates of ATE and ATT are moderately biased, classical estimates produce severe bias. Both strategies produce reasonable $ATUT$ estimates with the Bayesian estimation right on target. Finally, we consider one instrument.

9.7.3 One instrument

Suppose we have only one instrument, $Z_1+Z_2+Z_3+Z_4$. IV data augmented Gibbs sampler results are tabulated below.

statistic	β_1	β_0	θ_1
mean	12.08	13.95	−0.019
median	12.09	13.95	−0.019
standard dev	0.166	0.323	0.013
quantiles:			
minimum	11.42	12.95	−0.074
0.01	11.69	13.27	−0.051
0.025	11.75	13.35	−0.046
0.05	11.81	13.43	−0.041
0.10	11.87	13.53	−0.036
0.25	11.97	13.73	−0.027
0.75	12.19	14.18	−0.010
0.90	12.29	14.38	−0.002
0.95	12.35	14.50	0.003
0.975	12.40	14.59	0.006
0.99	12.47	14.69	0.011
maximum	12.67	15.12	0.033

Sample statistics for the parameters of the IV data augmented Gibbs sampler with one instrument applied to the treatment effect example

Selection bias emerges as β_0 again diverges from 12. This suggests marginal and average treatment effects are likely to be confounded. Sample statistics for average treatment effects as well as correlations, $\rho_{D,1}$, $\rho_{D,0}$, and $\rho_{1,0}$ are tabulated below.

statistic	ATE	ATT	$ATUT$	$\rho_{D,1}$	$\rho_{D,0}$	$\rho_{1,0}$
mean	-1.293	1.413	-4.000	0.797	-0.039	-0.048
median	-1.297	1.406	-4.000	0.801	-0.051	-0.061
standard dev	0.219	0.438	0.000	0.039	0.298	0.336
quantiles:						
minimum	-2.105	-0.211	-4.000	0.576	-0.757	-0.817
0.01	-1.806	0.389	-4.000	0.691	-0.615	-0.682
0.025	-1.738	0.525	-4.000	0.710	-0.554	-0.624
0.05	-1.665	0.670	-4.000	0.727	-0.503	-0.571
0.10	-1.572	0.855	-4.000	0.746	-0.429	-0.490
0.25	-1.435	1.130	-4.000	0.774	-0.272	-0.310
0.75	-1.147	1.705	-4.000	0.824	0.187	0.213
0.90	-1.005	1.989	-4.000	0.843	0.370	0.415
0.95	-0.930	2.141	-4.000	0.853	0.461	0.518
0.975	-0.861	2.277	-4.000	0.861	0.526	0.581
0.99	-0.795	2.409	-4.000	0.870	0.591	0.651
maximum	-0.625	2.750	-4.000	0.894	0.747	0.800

Sample statistics for average treatment effects and error correlations of the IV data augmented Gibbs sampler with one instrument applied to the treatment effect example

Weighted MTE estimates of average treatment effects are similar.

$estATE\,(MTE)$	$estATT\,(MTE)$	$estATUT\,(MTE)$
-1.957	0.060	-3.975

$ATUT$ is effectively estimated but the other average treatment effects are biased.

Classical results based on Heckman's inverse Mills control function strategy with one instrument are reported below for comparison. The selection equation estimated via probit is

$$\Pr(D \mid Z) = \Phi\left(-0.017\left(Z_1 + Z_2 + Z_3 + Z_4\right)\right) \quad PseudoR^2 = 0.009$$

The estimated outcome equations are

$$E[Y \mid X] = 14.000\,(1 - D) + 11.885D + NA\,(1 - D)\,\lambda_0 + 2.671D\lambda_1$$

and estimated average treatment effects are

$estATE$	$estATT$	$estATUT$
-2.115	NA	NA

While the Bayesian estimates of ATE and ATT are biased, the classical strategy fails to generate estimates for ATT and $ATUT$ — it involves a singular X matrix as there is no variation in λ_0.

9.8 A more standard example

Of course, the above sparse data example is an extreme case. By sparse we mean that even if there are a large number of draws, the draws cover a very sparse range of the sample space — in other words, there are a few draws potentially repeated a large number of times. In a setting where a large sample covers a broad range of the sample space, satisfaction of the instrumental variable condition (independence of the outcome errors) is satisfied via random draws. We next illustrate a protypical case with a simple example.[6]

A decision maker faces a binary choice where the latent choice equation (based on expected utility, EU, maximization) is

$$\begin{aligned} EU &= \gamma_0 + \gamma_1 x + \gamma_2 z + V \\ &= -1 + x + z + V \end{aligned}$$

x is an observed covariate, z is an observed instrument (both x and z have mean 0.5), and V is unobservable (to the analyst) contributions to expected utility. The outcome equations are

$$\begin{aligned} Y_1 &= \beta_0^1 + \beta_1^1 x + U_1 \\ &= 2 + 10x + U_1 \\ Y_0 &= \beta_0^0 + \beta_1^0 x + U_0 \\ &= 1 + 2x + U_0 \end{aligned}$$

Unobservables $\begin{bmatrix} V & U_1 & U_0 \end{bmatrix}^T$ are jointly normally distributed with expected value $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$ and variance $\Sigma = \begin{bmatrix} 1 & 0.7 & -0.7 \\ 0.7 & 1 & -0.1 \\ -0.7 & -0.1 & 1 \end{bmatrix}$. Clearly, the average treatment effect is

$$ATE = (2 + 10 * 0.5) - (1 + 2 * 0.5) = 5.$$

Even though OLS estimates the same quantity as ATE,

$$OLS = E\left[Y_1 \mid D = 1\right] - E\left[Y_0 \mid D = 0\right] = 7.56 - 2.56 = 5$$

selection is inherently endogenous. Further, outcomes are heterogeneous as[7]

$$ATT = E\left[Y_1 \mid D = 1\right] - E\left[Y_0 \mid D = 1\right] = 7.56 - 1.44 = 6.12$$

[6] This example is borrowed from Schroeder [2010], chapter 12.

[7] We can connect the dots by noting the average of the inverse Mills ratio is approximately 0.8 and recalling

$$\begin{aligned} ATE &= \Pr(D = 1)\, ATT + \Pr(D = 0)\, ATUT \\ &= 0.5\,(6.12) + 0.5\,(3.88) = 5 \end{aligned}$$

and

$$ATUT = E\left[Y_1 \mid D=0\right] - E\left[Y_0 \mid D=0\right] = 6.44 - 2.56 = 3.88$$

9.8.1 Simulation

To illustrate we generate 20 samples of $5,000$ observations each. For the simulation, x and z are independent and uniformly distributed over the interval $(0,1)$, and $\left[\begin{array}{ccc} V & U_1 & U_0 \end{array}\right]$ are drawn from a joint normal distribution with zero mean and variance Σ. If $EU_j > 0$, then $D_j = 1$, otherwise $D_j = 0$. Relatively diffuse priors are employed with mean zero and variance $100I$ for the parameters $\left[\begin{array}{ccc} \beta^1 & \beta^0 & \gamma \end{array}\right]$ and trivariate error $\left[\begin{array}{ccc} V & U_1 & U_0 \end{array}\right]$ distribution degrees of freedom parameter $\rho = 12$ and sums of squares variation ρI.[8] Data augmentation produces missing data for the latent choice variable EU plus counterfactuals $(Y_1 \mid D=0)$ and $(Y_0 \mid D=1)$.[9] Data augmentation permits collection of statistical evidence directly on the treatment effects. The following treatment effect statistics are collected:

$$\begin{aligned}
estATE &= \frac{1}{n}\sum_{j=1}^{n}\left(Y_{1j}^{*} - Y_{0j}^{*}\right) \\
estATT &= \frac{\sum_{j=1}^{n} D_j \left(Y_{1j}^{*} - Y_{0j}^{*}\right)}{\sum_{j=1}^{n} D_j} \\
estATUT &= \frac{\sum_{j=1}^{n} \left(1 - D_j\right)\left(Y_{1j}^{*} - Y_{0j}^{*}\right)}{\sum_{j=1}^{n} \left(1 - D_j\right)}
\end{aligned}$$

where Y_j^* is the augmented response. That is,

$$Y_{1j}^{*} = D_j Y_1 + \left(1 - D_j\right)\left(Y_1 \mid D=0\right)$$

and

$$Y_{0j}^{*} = D_j\left(Y_0 \mid D=1\right) + \left(1 - D_j\right) Y_0$$

[8] Initialization of the trivariate variance matrix for the Gibbs sampler is set equal to $100I$. Burn-in takes care of initialization error.

[9] Informativeness of the priors for the trivariate error variance is controlled by ρ. If ρ is small compared to the number of observations in the sample, the likelihood dominates the data augmentation.

9.8.2 Bayesian data augmentation and MTE

With a strong instrument in hand, this is an attractive setting to discuss a version of Bayesian data augmentation-based estimation of marginal treatment effects (MTE). As data augmentation generates repeated draws for unobservables V_j, $(Y_{1j} \mid D_j = 0)$, and $(Y_{0j} \mid D_j = 1)$, we exploit repeated samples to describe the distribution for $MTE\left(u_D\right)$ where V is transformed to uniform $(0,1)$, $u_D = p_v$. For each draw, $V = v$, we determine $u_D = \Phi\left(v\right)$ and calculate $MTE\left(u_D\right) = E\left[Y_1 - Y_0 \mid u_D\right]$.

MTE is connected to standard population-level treatment effects, ATE, ATT, and $ATUT$, via non-negative weights whose sum is one

$$\begin{aligned} w_{ATE}\left(u_D\right) &= \frac{\sum_{j=1}^{n} I\left(u_D\right)}{n} \\ w_{ATT}\left(u_D\right) &= \frac{\sum_{j=1}^{n} I\left(u_D\right) D_j}{\sum_{j=1}^{n} D_j} \\ w_{ATUT}\left(u_D\right) &= \frac{\sum_{j=1}^{n} I\left(u_D\right)\left(1 - D_j\right)}{\sum_{j=1}^{n}\left(1 - D_j\right)} \end{aligned}$$

where probabilities p_k refer to bins from 0 to 1 by increments of 0.01 for indicator variable

$$\begin{array}{ll} I\left(u_D\right) = 1 & u_D = p_k \\ I\left(u_D\right) = 0 & u_D \neq p_k \end{array}$$

Simulation results

Since the Gibbs sampler requires a burn-in period for convergence, for each sample we take $4,000$ conditional posterior draws, treat the first $3,000$ as the burn-in period, and retain the final $1,000$ draws for each sample, in other words, a total of $20,000$ draws are retained. Parameter estimates for the simulation are reported in the table below.

statistic	β_0^1	β_1^1	β_0^0	β_1^0
mean	2.118	9.915	1.061	2.064
median	2.126	9.908	1.059	2.061
std.dev.	0.100	0.112	0.063	0.102
minimum	1.709	9.577	0.804	1.712
maximum	2.617	10.283	1.257	2.432
statistic	γ_0	γ_1	γ_2	
mean	-1.027	1.001	1.061	
median	-1.025	0.998	1.061	
std.dev.	0.066	0.091	0.079	
minimum	-1.273	0.681	0.729	
maximum	-0.783	1.364	1.362	
statistic	$cor\,(V, U_1)$	$cor\,(V, U_0)$	$cor\,(U_1, U_0)$	
mean	0.621	-0.604	-0.479	
median	0.626	-0.609	-0.481	
std.dev.	0.056	0.069	0.104	
minimum	0.365	-0.773	-0.747	
maximum	0.770	-0.319	0.082	

$$Y_1 = \beta_0^1 + \beta_1^1 x + U_1$$
$$Y_0 = \beta_0^0 + \beta_1^0 x + U_0$$
$$EU = \gamma_0 + \gamma_1 x + \gamma_2 z + V$$

McMC parameter estimates for prototypical example

McMC estimated average treatment effects are reported in the table below

statistic	*estATE*	*estATT*	*estATUT*
mean	4.992	6.335	3.635
median	4.996	6.329	3.635
std.dev.	0.087	0.139	0.117
minimum	4.703	5.891	3.209
maximum	5.255	6.797	4.067

McMC estimates of average treatment effects for prototypical example

and sample statistics are reported in the table below.

statistic	*ATE*	*ATT*	*ATUT*	*OLS*
mean	5.011	6.527	3.481	5.740
median	5.015	6.517	3.489	5.726
std.dev.	0.032	0.049	0.042	0.066
minimum	4.947	6.462	3.368	5.607
maximum	5.088	6.637	3.546	5.850

McMC average treatment effect sample statistics for prototypical example

The treatment effect estimates are consistent with their sample statistics despite the fact that bounding the unidentified correlation between U_1 and U_0 produces a rather poor estimate of this parameter.

In addition, we report results on marginal treatment effects. The table below reports simulation statistics from weighted averages of MTE employed to recover standard population-level treatment effects, ATE, ATT, and $ATUT$.

statistic	$estATE$	$estATT$	$estATUT$
mean	4.992	5.861	4.114
median	4.980	5.841	4.115
std.dev.	0.063	0.088	0.070
minimum	4.871	5.693	3.974
maximum	5.089	6.003	4.242

McMC MTE-weighted average treatment effects for prototypical example

Nonconstancy of $MTE\left(u_D\right)$ along with marked differences in $estATE$, $estATT$, and $estATUT$ provide support for heterogeneous response. The MTE-weighted average treatment effect estimates are very comparable (perhaps slightly dampened) to the previous estimates and average treatment effect sample statistics.