Exploring Marginal Treatment Effects

Flexible estimation using Stata

Martin Eckhoff Andresen

Statistics Norway

Oslo, September 12th 2018

Motivation

- **Instrumental variables** (IV) estimators solve endogeneity problems
- When there is **heterogenous returns**, IV estimate LATE:
  - Average treatment effect among **compliers**
  - Not always of interest!
- **Marginal Treatment Effects** allows you to
  - Go **beyond LATE** in settings with essential heterogeneity
  - Capture the **full distribution** of treatment effects
  - Allow us to back out commonly used **treatment effect parameters**
  - Unify IV methods, **selection** models and **control function** approaches

This paper

- Presents the **theory** of Marginal Treatment Effects aimed at the **applied empiricist**
- Highlights similarities to **selection models** and **control function** approaches
- Introduces the new **Stata package** `mtefe` for estimating MTEs
- Performs Monte Carlo **simulations** to investigate the robustness of the estimators

A motivating example: College and wages

$$D = a + bX + cZ + \overbrace{d\text{ability} + eZ \times \text{ability} + \mu}^{\text{unobserved}}$$
$$w = f + gX + hD \; + \underbrace{i\text{ability} + jD \times \text{ability} + \epsilon}_{\text{unobserved}}$$

- If $d \neq 0 \neq i$: Selection problem
- If $j = 0$: IV recovers the ATE with a valid instrument Z
- If $j \neq 0$: IV recovers a **local average treatment effect**
- Relative size of *LATE* vs. *ATE* depends on
  - what individuals are shifted into treatment by the instrument - e
  - what individuals have higher or lower treatment effects - j

A generalized Roy model

$$Y_j = \mu_j(X) + U_j \qquad \text{for } j = 0, 1 \qquad (1)$$

$$Y = DY_1 + (1 - D)Y_0 \qquad (2)$$

$$D = \mathbb{1}\left[Z\gamma > V\right] \qquad \text{where } Z = X, Z_- \qquad (3)$$

- Without loss of generality normalize the scale of V
  - $D = 1 \Leftrightarrow \gamma Z > V \Leftrightarrow F_V(Z\gamma) > F_V(V) \Leftrightarrow P(Z) > U_D$
  - $U_D \sim U(0, 1)$: Percentiles of the **unobserved resistance**
- Treatment effect: $\boldsymbol{\beta} = Y_1 - Y_0 = \mu_1(X) - \mu_0(X) + U_1 - U_0$
- With **essential heterogeneity**: Sorting on unobserved gains
  - $\text{cov}(\boldsymbol{\beta}, D \mid X) \neq 0$
  - Treatment decision made with knowledge of unobserved gains

Marginal Treatment Effects

$$\begin{aligned} \text{MTE}(x,u) &\equiv \mathbb{E}(Y_1 - Y_0 | U_d = u, X = x) \\ &= \mu_1(x) - \mu_0(x) + \mathbb{E}(U_1 - U_0 \mid U_D = u) \end{aligned}$$

Average β for people with a **particular distaste for treatment** and x

Björklund and Moffitt (1987), Heckman and coauthors (1997; 1999; 2005; 2007), Cornelissen et al. (2016); Brinch et al. (2015).

LATE vs MTE

With two particular values of an instrument, z and z', the Wald estimator is

$$\text{LATE}(x) = \frac{\mathbb{E}(Y|X=x, Z_{-}=z') - \mathbb{E}(Y|X=x, Z_{-}=z)}{\mathbb{E}(D|X=x, Z_{-}=z') - \mathbb{E}(D|X=x, Z_{-}=z)}$$

This is a **Local Average Treatment Effect**

- People who choose treatment when $Z_{-} = z'$, but not when $Z_{-} = z$
- In the choice model: People with $P(x, z') < U_D \leq P(x, z)$:

$$\text{LATE}(x, z, z') = \mu_1(x) - \mu_0(x) + \mathbb{E}(U_1 - U_0 | P(x, z') < U_D \leq P(x, z))$$
$$\text{MTE}(x, u) = \mu_1(x) - \mu_0(x) + \mathbb{E}(U_1 - U_0 | U_D = u)$$

- MTE is a **limit form of LATE** (Heckman, 1997)

Back to the motivating example

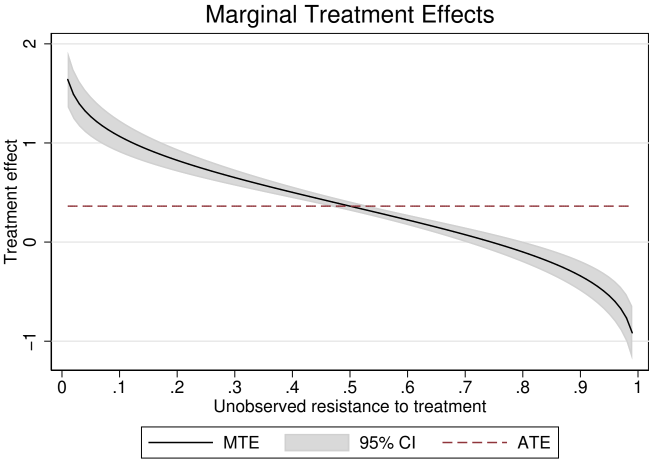
$$D = a + bX + cZ + \overbrace{d\text{ability} + eZ \times \text{ability} + \mu}^{\text{unobserved}}$$
$$w = f + gX + hD \;\; + \underbrace{i\text{ability} + jD \times \text{ability} + \epsilon}_{\text{unobserved}}$$

In the choice model, every omitted variable will enter U_0, U_1, U_D.

- High-ability people will have lower U_D if $d > 0$
- ...and **higher unobserved treatment effects** $(U_1 - U_0)$ if $j > 0$
- Should lead to a downward sloping MTE - $\text{cov}(U_1 - U_0, U_D) < 0$

This selection pattern is precisely what MTE estimates

An example MTE curve



Standard IV assumptions

Interpreting IV as **LATE** (Imbens and Angrist, 1994) requires:

Exclusion $Y_j \perp Z_{-}|X$. The instrument affect outcomes only through the probability of treatment $| X$

Relevance $P(z) \neq P(z')$. Treatment is a nontrivial function of the instrument

Monotonicity $P(z) \geq P(z') \; \forall i$ two values of the instrument cannot shift some people in and others out

- Monotonicity should hold between all possible pairs z, z'
- These assumptions imply and are implied by the model in Eq. 1-3 (Vytlacil, 2002)

The separability assumption

Best case scenario: Estimate MTEs with no more assumptions than IV

- Estimate MTE within each cell of X, aggregate
- In practice: Limited data and support. Instead assume

Separability $\mathbb{E}(U_j \mid X, U_D) = \mathbb{E}(U_j \mid U_D)$

Implied by, but weaker than, full independence

- All X do is shift the MTE curve up or down
- Same assumption as in selection models

Usually also work with linear version of $\mu_j(x) = x\beta_j$

Estimation methods

- First estimate the propensity scores $P(Z)$
- Local Instrumental Variables
  - The derivative of the conditional expectation of Y wrt. p
  - $\text{MTE}(x, u) = \frac{\partial \mathbb{E}(Y|x,p)}{\partial p} |_{u=p}$
- Separate approach
  - Estimate outcome given x, p separately controlling selection
  - $\text{MTE}(x, u) = \mathbb{E}(Y_1 \mid x, U_D = u) - E(Y_0 \mid x, U_D = u)$
  - Control selection via control function - similar to selection models
- Maximum likelihood (joint normal model only)

Functional forms

- $(U_{0,} U_1, V)$ joint normal: Heckman selection
- $\mathbb{E}(U_j \mid U_D = u) = \sum_1^K \pi_k (u^k - \frac{1}{k+1})$: polynomial model
- Polynomial model with splines
- Semiparametric model
  - Estimate partial linear model of $\mathbb{E}(Y \mid X, p) = X\beta_0 + X(\beta_1 - \beta_0)p + K(p)$
  - Using double residual regression (Robinson, 1988)

The `mtefe` package

- Acceps fixed effects in all independent varlists
- Supports weights (pweights, fweights)
- Supports Local IV, separate approach and maximum likelihood estimation
- More flexible MTE models, including spline functions
- Calculates treatment effect parameters from results
- Analytic standard errors and bootstrap including first stage
- Improved graphical output (`mtefeplot`)

Brave and Walstrum (2014)

The `mtefe` command

`mtefe` *depvar* [*indepvars*] (*depvar*$_t$ = *varlist*$_{iv}$) [*if*] [*in*] [*weight*] [,
`polynomial(#)` `splines(`*numlist*`)` `semiparametric` `restricted(`*varlist*$_r$`)`
`separate` `mlikelihood` `link(`*string*`)` `+ other options`]

- Follows Stata's IV syntax
- Accepts fixed effects (i.varname)
- Several options follow similar options in `margte`

Example output I

```
. mtefe_gendata, obs(10000) districts(10)
.
. mtefe lwage exp exp2 i.district (col=distCol)
Parametric normal MTE model                          Observations : 10000
Treatment model: Probit
Estimation method: Local IV
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beta0						
exp	.0358398	.0064408	5.56	0.000	.0232145	.0484651
exp2	-.0008453	.0002019	-4.19	0.000	-.0012411	-.0004496
district						
2	.2352456	.0680412	3.46	0.001	.1018712	.36862
3	.6294914	.0701091	8.98	0.000	.4920634	.7669194
4	.0131179	.0597721	0.22	0.826	-.1040474	.1302832
5	.0338606	.0705835	0.48	0.631	-.1044974	.1722186
6	.1699366	.0605086	2.81	0.005	.0513275	.2885458
7	-.1899241	.060115	-3.16	0.002	-.3077617	-.0720865
8	-.1842254	.0676843	-2.72	0.007	-.3169003	-.0515504
9	-.7908301	.0578436	-13.67	0.000	-.9042153	-.677445
10	-.4432749	.0597237	-7.42	0.000	-.5603455	-.3262044
_cons	3.164706	.0650331	48.66	0.000	3.037228	3.292184
beta1-beta0						

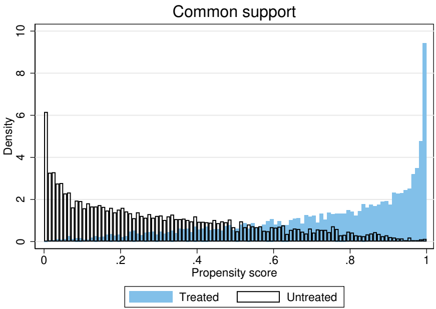
Example output II

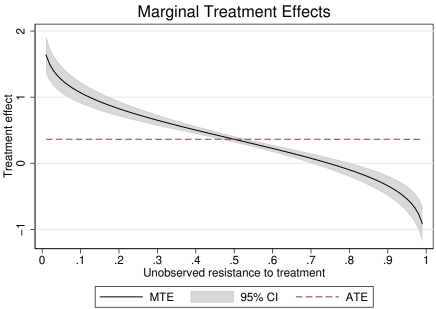
exp	-.0386384	.010241	-3.77	0.000	-.0587128	-.018564
exp2	.0012967	.0003288	3.94	0.000	.0006523	.0019412
district						
2	.265112	.107039	2.48	0.013	.0552939	.4749301
(output omitted)						
10	.3143661	.1072555	2.93	0.003	.1041237	.5246085
_cons	.4255863	.0983572	4.33	0.000	.2327863	.6183863
k						
mills	-.4790282	.0611081	-7.84	0.000	-.5988124	-.359244
effects						
ate	.3283373	.0242932	13.52	0.000	.2807177	.3759568
att	.5369432	.0388809	13.81	0.000	.4607287	.6131576
atut	.1195067	.0384691	3.11	0.002	.0440995	.194914
late	.3279726	.0245142	13.38	0.000	.2799198	.3760254
mprte1	.3463148	.0256971	13.48	0.000	.2959433	.3966862
mprte2	.3309428	.024298	13.62	0.000	.2833137	.3785719
mprte3	-.016257	.0498984	-0.33	0.745	-.1140679	.0815538

```
Test of observable heterogeneity, p-value                   0.0000
Test of essential heterogeneity, p-value                    0.0000

Note: Analytical standard errors ignore the facts that the propensity score,
```
(output omitted)

Marginal Treatment Effects of college





Interpreting heterogeneity

- The unobserved dimension
    - **Depends on what is observed**!
    - Positive selection on unobserved gains: MTE is downward sloping
        - In line with predictions from a simple Roy model
        - Consistent with treatment decisions made with **knowledge of unobserved gains**
- The observed dimension
    - Positive selection if $\gamma \times (\beta_1 - \beta_0) > 0$
        - X that leads to more treatment also leads to higher treatment effects
    - Negative selection if $\gamma \times (\beta_1 - \beta_0) < 0$

MTEs unify treatment effect parameters

Using MTE(x, u), we can calculate any treatment effect parameter as

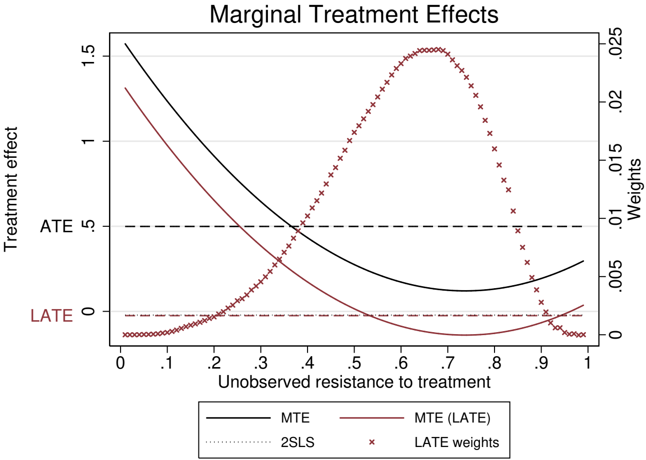
$$\int_0^1 \omega(u)\text{MTE}(\bar{x}, u)du = \bar{x}(\beta_1 - \beta_0) + \int_0^1 \omega(u)k(u)du$$

- $\omega(u)$ is the density of U_D in the population of interest
- $\bar{x}$ is the average x in the population of interest

Where the population of interest depends on the parameter:

- ATE: Everyone, $\omega(u) = 1$, $\bar{x}$ is average x
- ATT: Population has $D = 1 \Leftrightarrow U_D \leq p$, $\bar{x}$ is average among treated
- ATUT: Population has $D = 0 \Leftrightarrow U_D > p$, $\bar{x}$ is average among untreated
- LATE/IV: Population is compliers
- PRTE/MPRTE: Population is people shifted by policy

Local Average Treatment Effects



Practical advice for users

- Interpret the unobserved dimension **in light of observables**
- Know your setting, argue explicitly for what U_D could pick up
- Use this to defend the separability assumption
- Use **semiparametric methods** to guide your choice of functional form
- Show robustness to
  - Choice of **functional form**
  - Use of **estimation method**

Conclusion

- Marginal treatment effects should be in your toolbox
- **Heterogeneous returns** is the more reasonable baseline case
- MTE analysis estimate the full **distribution of treatment effects** and thus go beyond LATE
  - But usually at the cost of **stricter assumptions**
  - Unless you have an instrument that work without covariates and generate full support
- ...but MTE aren't all that new - closely related to selection models.
- The `mtefe` package does the work for you (please report bugs)

References I

Björklund, A. and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-selection," *The Review of Economics and Statistics*, 69, 42–49.

Brave, S. and T. Walstrum (2014): "Estimating marginal treatment effects using parametric and semiparametric methods," *Stata Journal*, 14, 191–217(27).

Brinch, C., M. Mogstad, and M. Wiswall (2015): "Beyond LATE with a discrete instrument," Forthcoming in Journal of Political Economy.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016): "From LATE to MTE: Alternative methods for the evaluation of policy interventions," *Labour Economics*, 41, 47 – 60, sOLE/EALE conference issue 2015.

Heckman, J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462.

Heckman, J. J. and E. J. Vytlacil (1999): "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), 4730–4734.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

References II

——— (2007): "Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," Elsevier, vol. 6, Part B of *Handbook of Econometrics*, 4875 – 5143.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, pp. 467–475.

Robinson, P. (1988): "Root- N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–54.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.