Technical Appendix for 'Simple Estimators for Treatment Parameters in a Latent Variable Framework'

J. Heckman, J.L. Tobias and E. Vytlacil[1]

Let $\lambda_1(t) := \frac{\phi(t)}{\Phi(t)}$, $\lambda_0(t) := \frac{\phi(-t)}{\Phi(-t)} = \frac{\phi(t)}{1-\Phi(t)}$, $\alpha^1 = \rho_1\sigma_1$, $\alpha^0 = -\rho_0\sigma_0$, $W := (D, Y, X, Z)$, $\tilde{\beta}_j := (\beta^j, \alpha^j)$, and $\tilde{\beta} := (\beta^0, \beta^1, \alpha^0, \alpha^1)$.

We use a standard estimation process where the first step uses maximum likelihood to estimate θ from the appropriate binary choice model. The second step uses this estimate and does OLS regression of Y on X and the correction term. The treatment parameters are then estimated by plugging in estimated parameters from the two-step procedure into the expressions provided above.

We now derive the asymptotic distribution theory for the resulting estimators of the treatment parameters. We follow Newey and McFadden (1994) in viewing a multi-step estimator as a GMM estimator for the stacked moment conditions with the identity matrix as the weighting matrix. For the treatment parameters that condition on covariates, the asymptotic distribution of the plug-in estimators for these parameters is derived by first deriving the asymptotic distribution theory for the two-step estimator and then using that each plug-in estimator for a treatment parameters is simply a linear or smooth nonlinear function of the parameters estimated by the two-step procedure. In the case of treatment parameters that average over covariates, we consider the resulting estimator as a three-step estimator where the averaging over covariates is the third-step. We derive results for the case of joint normality. Our results can be easily modified to provide the asymptotic distribution theory corresponding to the estimators of the other models discussed in the text.

We first derive the asymptotic distribution theory for the two-step selection-corrected estimator. In order to derive the distribution theory for our parameters, we need not only the asymptotic distribution theory for the estimator of the second stage parameters, but also the covariance between the estimator for the θ coefficient and the estimator for the second stage parameters. We thus first derive the asymptotic distribution theory for the two-stage estimator. We then use these results in section 3 to derive the asymptotic distribution theory for the plug-in estimators of treatment parameters that condition on covariates, and in section 4 to derive the asymptotic distribution theory for the plug-in estimators of treatment parameters that do not condition on covariates.

We can view the two-step estimator as a GMM estimator with the identity matrix as the weighting matrix. Under standard regularity conditions, following the arguments in Newey

[1]The assistance of Nese Yildiz is gratefully acknowledged.

and McFadden (Theorem 6.1), we have

$$\sqrt{N}\left(\begin{bmatrix}\hat{\beta}^1\\ \hat{\alpha}^1\\ \hat{\beta}^0\\ \hat{\alpha}^0\\ \hat{\theta}\end{bmatrix}-\begin{bmatrix}\beta^1\\ \alpha^1\\ \beta^0\\ \alpha^0\\ \theta\end{bmatrix}\right)\to N(0,V)$$

where $V = G^{-1}\Omega G^{-1\prime}$.

1 Computing V:

To compute V, we need to compute G and Ω. For this purpose we define

$$\begin{aligned} g_1(W,\theta,\tilde{\beta}) &:= D\begin{bmatrix}X\\ \lambda_1(Z'\theta)\end{bmatrix}[Y - X'\beta^1 - \alpha^1\lambda_1(Z'\theta)] \\ g_0(W,\theta,\tilde{\beta}) &:= (1-D)\begin{bmatrix}X\\ \lambda_0(Z'\theta)\end{bmatrix}[Y - X'\beta^0 - \alpha^0\lambda_0(Z'\theta)] \\ m(W,\theta) &:= Z\frac{\phi(Z'\theta)}{\Phi(Z'\theta)}\frac{D-\Phi(Z'\theta)}{1-\Phi(Z'\theta)} = Z[D\lambda_1(Z'\theta) - (1-D)\lambda_0(Z'\theta)] \end{aligned}$$

Now we let $g(W) := \begin{pmatrix} g_1(W,\theta,\tilde{\beta})\\ g_0(W,\theta,\tilde{\beta})\\ m(W,\theta)\end{pmatrix}$. Then define $G := E\left[\frac{\partial g(W)}{\partial(\tilde{\beta}_1',\tilde{\beta}_0',\theta')}\right]$. Alternatively, we could write $G = \begin{bmatrix} G_{1,\tilde{\beta}_1} & 0 & G_{1,\theta}\\ 0 & G_{0,\tilde{\beta}_0} & G_{0,\theta}\\ 0 & 0 & M\end{bmatrix} = \begin{bmatrix} E\left(\frac{\partial g_1}{\partial\tilde{\beta}_1}\right) & 0 & E\left(\frac{\partial g_1}{\partial\theta}\right)\\ 0 & E\left(\frac{\partial g_0}{\partial\tilde{\beta}_0}\right) & E\left(\frac{\partial g_0}{\partial\theta}\right)\\ 0 & 0 & E\left(\frac{\partial m}{\partial\theta}\right)\end{bmatrix}$, where

$G_{1,\tilde{\beta}_1} = -\begin{bmatrix} E(\Phi(Z'\theta)XX') & E(\phi(Z'\theta)X)\\ E(\phi(Z'\theta)X') & E(\phi(Z'\theta)\lambda_1(Z'\theta))\end{bmatrix}$, $G_{0,\tilde{\beta}_0} = -\begin{bmatrix} E(\Phi(-Z'\theta)XX') & E(\phi(Z'\theta)X)\\ E(\phi(Z'\theta)X') & E(\phi(Z'\theta)\lambda_0(Z'\theta))\end{bmatrix}$

$G_{1,\theta} = \begin{bmatrix} -\alpha^1 E(DXZ'\lambda_1'(Z'\theta))\\ E(DZ'\lambda_1'(Z'\theta)[Y - X'\beta^1 - 2\alpha^1\lambda_1(Z'\theta))]\end{bmatrix} = \begin{bmatrix} E(-\alpha^1 XZ'\lambda_1'(Z'\theta)\Phi(Z'\theta))\\ E(-Z'\lambda_1'(Z'\theta)\alpha^1\phi(Z'\theta)\end{bmatrix}$ for

$E(-\alpha^1 DXZ'\lambda_1'(Z'\theta)) = E_{X,Z}[-\alpha^1 XZ\lambda_1'(Z'\theta)E(D|X,Z)] = E(-\alpha^1 XZ\lambda_1'(Z'\theta)\Phi(Z'\theta))$

$E(DZ'\lambda_1'(Z'\theta)[Y - X'\beta^1 - 2\alpha^1\lambda_1(Z'\theta)]) = E(Z'\lambda_1'(Z'\theta)(DU^1 - 2\alpha^1\lambda_1(Z'\theta)D))$

$= E_{X,Z}[Z'\lambda_1'(Z'\theta)(E(DU^1|X,Z) - 2\alpha^1\lambda_1(Z'\theta)E(D|X,Z))] = E(-Z'\lambda_1'(Z'\theta)\alpha^1\phi(Z'\theta))$.

Similar calculations yield:

$G_{0,\theta} = \begin{bmatrix} -E((1-D)\alpha^0 XZ'\lambda_0'(Z'\theta))\\ E((1-D)Z'\lambda_0'(Z'\theta)[Y - X'\beta^0 - 2\alpha^0\lambda_0(Z'\theta))]\end{bmatrix} = \begin{bmatrix} E(-\alpha^0 XZ'\Phi(-Z'\theta)\lambda_0'(Z'\theta))\\ E(-Z'\alpha^0\phi(Z'\theta)\lambda_0'(Z'\theta))\end{bmatrix}$

Finally, $M = E[ZZ'(D\lambda_1'(Z'\theta) - (1-D)\lambda_0'(Z'\theta))] = -E\left[ZZ'\frac{\phi(Z'\theta)^2}{\Phi(Z'\theta)(1-\Phi(Z'\theta))}\right]$

Now we are ready to calculate G^{-1}:

$$G^{-1} = \begin{bmatrix} G^{-1}_{1,\tilde{\beta}_1} & 0 & -G^{-1}_{1,\tilde{\beta}_1} G_{1,\theta} M^{-1} \\ 0 & G^{-1}_{0,\tilde{\beta}_0} & -G^{-1}_{0,\tilde{\beta}_0} G_{0,\theta} M^{-1} \\ 0 & 0 & M^{-1} \end{bmatrix}$$

$G^{-1}_{1,\tilde{\beta}_1} = -\begin{bmatrix} \Gamma^{11}_{1,\tilde{\beta}_1} & \Gamma^{12}_{1,\tilde{\beta}_1} \\ \Gamma^{21}_{1,\tilde{\beta}_1} & \Gamma^{22}_{1,\tilde{\beta}_1} \end{bmatrix}$, $G^{-1}_{0,\tilde{\beta}_0} = -\begin{bmatrix} \Gamma^{11}_{0,\tilde{\beta}_0} & \Gamma^{12}_{0,\tilde{\beta}_0} \\ \Gamma^{21}_{0,\tilde{\beta}_0} & \Gamma^{22}_{0,\tilde{\beta}_0} \end{bmatrix}$ where

$\Gamma^{22}_{1,\tilde{\beta}_1} = \frac{1}{E(\phi(Z'\theta)\lambda_1(Z'\theta)) - E(\phi(Z'\theta)X')[E(\Phi(Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)}$

$\Gamma^{21}_{1,\tilde{\beta}_1} = -\Gamma^{22}_{1,\tilde{\beta}_1} E(\phi(Z'\theta)X')[E(\Phi(Z'\theta)XX')]^{-1}$, $\Gamma^{12}_{1,\tilde{\beta}_1} = -[E(\Phi(Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)\Gamma^{22}_{1,\tilde{\beta}_1}$

$\Gamma^{11}_{1,\tilde{\beta}_1} = [E(\Phi(Z'\theta)XX')]^{-1} + [E(\Phi(Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)\Gamma^{22}_{1,\tilde{\beta}_1}E(\phi(Z'\theta)X')[E(\Phi(Z'\theta)XX')]^{-1}$

$\Gamma^{22}_{0,\tilde{\beta}_0} = \frac{1}{E(\phi(Z'\theta)\lambda_0(Z'\theta)) - E(\phi(Z'\theta)X')[E(\Phi(-Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)}$

$\Gamma^{21}_{0,\tilde{\beta}_0} = -\Gamma^{22}_{0,\tilde{\beta}_0} E(\phi(Z'\theta)X')[E(\Phi(-Z'\theta)XX')]^{-1}$, $\Gamma^{12}_{0,\tilde{\beta}_0} = -[E(\Phi(-Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)\Gamma^{22}_{0,\tilde{\beta}_0}$

$\Gamma^{11}_{0,\tilde{\beta}_0} = [E(\Phi(-Z'\theta)XX')]^{-1} + [E(\Phi(-Z'\theta)XX')]^{-1}E(\phi(Z'\theta)X)\Gamma^{22}_{0,\tilde{\beta}_0}E(\phi(Z'\theta)X')[E(\Phi(-Z'\theta)XX')]^{-1}$

On the other hand, $\Omega := E[g(W)g(W)']$. Since $D(1-D) = 0$, $g_1(W)g_0(W) = 0$. Moreover, $E(g_1(W)|X,Z,D) = E(g_0(W)|X,Z,D) = 0$ so that $E(m(W)g_1(W)) = E(m(W)g_0(W) = 0$. Therefore,

$$\Omega = E\left(\begin{bmatrix} g_1(W)g_1(W)' & 0 & 0 \\ 0 & g_0(W)g_0(W)' & 0 \\ 0 & 0 & m(W)m(W)' \end{bmatrix}\right)$$

Let $\Psi(W) := M^{-1}m(W)$.

$$Asymp.\ V(\hat{\tilde{\beta}}^1) = G^{-1}_{1,\tilde{\beta}^1} E\big(g_1(W)g_1(W)' + G_{1,\theta}\Psi(W)\Psi(W)'G'_{1,\theta}\big)G^{-1\prime}_{1,\tilde{\beta}^1}$$

$$Asymp.\ V(\hat{\tilde{\beta}}^0) = G^{-1}_{0,\tilde{\beta}^0} E\big(g_0(W)g_0(W)' + G_{0,\theta}\Psi(W)\Psi(W)'G'_{0,\theta}\big)G^{-1\prime}_{0,\tilde{\beta}^0}$$

$$Asymp.\ V(\hat{\theta}) = E(\Psi(W)\Psi(W)')$$

$$Asymp.\ Cov(\hat{\tilde{\beta}}^1, \hat{\tilde{\beta}}^0) = G^{-1}_{1,\tilde{\beta}^1} G_{1,\theta} E\big(\Psi(W)\Psi(W)'\big)G'_{0,\theta}G^{-1\prime}_{0,\tilde{\beta}^0}$$

$$Asymp.\ Cov(\hat{\tilde{\beta}}^1, \hat{\theta}) = -G^{-1}_{1,\tilde{\beta}^1} G_{1\theta} E\Big[\Psi(W)\Psi(W)'\Big]$$

$$Asymp.\ Cov(\hat{\tilde{\beta}}^0, \hat{\theta}) = -G^{-1}_{0,\tilde{\beta}^0} G_{0\theta} E\Big[\Psi(W)\Psi(W)'\Big]$$

2 Asymptotic Distribution of $\widehat{ATE}, \widehat{TT}, \widehat{LATE}$ and $\widehat{MTE}$ given x and z:

Estimator for Average Treatment Effect conditional on $X = x$:

$$\widehat{ATE}(x) \equiv x'(\hat{\beta}^1 - \hat{\beta}^0)$$

By theorems 6.1.3 and 6.1.4 of Amemiya (1994):

$$\sqrt{N}\left(\widehat{ATE}(x) - ATE(x)\right) \rightarrow N(0, x'(V_{11} - 2V_{13} + V_{33})x)$$

Estimator for the Effect of Treatment on the Treated conditional on $X = x, Z = z$ and $D(z) = 1$:

$$\widehat{TT}(x, z, D(z) = 1) \equiv x'(\hat{\beta}^1 - \hat{\beta}^0) + (\hat{\alpha}^1 + \hat{\alpha}^0)\frac{\phi(z'\hat{\theta})}{\Phi(z'\hat{\theta})}$$

A simple Taylor expansion and application of the same theorems yield:

$$\sqrt{N}\left(\widehat{TT}(x, z, D(z) = 1) - TT(x, z, D(z) = 1)\right) \rightarrow N(0, S_1 V S_1')$$

where $S_1 = \left[\begin{array}{cccccc} x' & \frac{\phi(z'\theta)}{\Phi(z'\theta)} & -x' & \frac{\phi(z'\theta)}{\Phi(z'\theta)} & (\alpha^1 + \alpha^0)\frac{\phi'(z'\theta)\Phi(z'\theta) - (\phi(z'\theta))^2}{(\Phi(z'\theta))^2} \end{array}\right]$.
Estimator for Local Average Treatment Effect conditional on x, $D(z) = 0$, $D(\tilde{z}) = 1$:

$$\widehat{LATE}(x, D(z) = 0, D(\tilde{z}) = 1) = x'(\hat{\beta}^1 - \hat{\beta}^0) + (\hat{\alpha}^1 + \hat{\alpha}^0)\frac{\phi(\tilde{z}'\hat{\theta}) - \phi(z'\hat{\theta})}{\Phi(\tilde{z}'\hat{\theta}) - \Phi(z'\hat{\theta})}$$

Following similar steps we get:

$$\sqrt{N}\left(\widehat{LATE}(x, D(z) = 0, D(\tilde{z}) = 1) - LATE(x, D(z) = 0, D(\tilde{z}) = 1)\right) \rightarrow N(0, S_2 V S_2')$$

$$S_2 = \left[\begin{array}{ccccc} x' & \frac{\phi(\tilde{z}'\theta) - \phi(z'\theta)}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)} & -x' & \frac{\phi(\tilde{z}'\theta) - \phi(z'\theta)}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)} & (\alpha^1 + \alpha^0)\left(\frac{(\tilde{z}'\phi'(\tilde{z}'\theta) - z'\phi'(z'\theta))}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)} - \frac{(\phi(\tilde{z}'\theta) - \phi(z'\theta))(\tilde{z}'\phi(\tilde{z}'\theta) - z'\phi(z'\theta))}{(\Phi(\tilde{z}'\theta) - \Phi(z'\theta))^2}\right) \end{array}\right]$$

Finally, estimator for Marginal Treatment Effect conditional on $X = x$, $U^D = u^D$:

$$\widehat{MTE}(x, u^D) = x'(\hat{\beta}^1 - \hat{\beta}^0) + (\hat{\alpha}^1 + \hat{\alpha}^0)u^D$$

Its asymptotic distribution is given by:

$$\begin{aligned}\sqrt{N}\left(\widehat{MTE}(x, u^D) - MTE(x, u^D)\right) \rightarrow N(0 \quad &, \quad x'(V_{11} - 2V_{13} + V_{33})x + u^{D\prime}(V_{22} + 2V_{24} \\ &+ \quad V_{44})u^D + 2x'(V_{12} - V_{32} + V_{14} - V_{34})u^D)\end{aligned}$$

3 Asymptotic Distribution of $\widehat{ATE}, \widehat{TT}, \widehat{LATE}$ and $\widehat{MTE}$:

Estimation and finding the asymptotic distribution of $\widehat{ATE}$, $\widehat{TT}$, $\widehat{LATE}$ and $\widehat{MTE}$ require a different moment condition in each case in addition to the three moment conditions above, because we have one more parameter to estimate. We let g_e denote that extra moment condition and γ denote the extra parameter in each case. Then under regularity conditions,

$$\sqrt{N}\left(\begin{bmatrix} \hat{\beta}^1 \\ \hat{\alpha}^1 \\ \hat{\beta}^0 \\ \hat{\alpha}^0 \\ \theta \\ \hat{\gamma} \end{bmatrix} - \begin{bmatrix} \beta^1 \\ \alpha^1 \\ \beta^0 \\ \alpha^0 \\ \theta \\ \gamma \end{bmatrix}\right) \to N(0, V_e)$$

where $V_e = G_e^{-1}\Omega_e G_e^{-1\prime}$, with $\Omega_e = E[g^e(W)g^e(W)']$ and $G_e = E\left[\frac{\partial g^e}{\partial(\tilde{\beta}^1, \tilde{\beta}^0, \theta, \gamma)}\right]$, and $g^e(W) := \begin{pmatrix} g_1(W)' & g_0(W)' & m(W)' & g_e(W) \end{pmatrix}$. Below, we will often have to calculate G_e^{-1}. For this purpose, let us express G_e^{-1} as $\begin{bmatrix} G & B \\ C & H \end{bmatrix}$ where G is as before. Then

$$G_e^{-1} = \begin{bmatrix} G^{-1} + G^{-1}B(H - CG^{-1}B)^{-1}CG^{-1} & -G^{-1}B(H - CG^{-1}B)^{-1} \\ -(H - CG^{-1}B)^{-1}CG^{-1} & (H - CG^{-1}B)^{-1} \end{bmatrix}$$

3.1 Asymptotic Distribution of $\widehat{ATE}$:

Let $\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n} x_i'(\hat{\beta}^1 - \hat{\beta}^0)$. Let, $g_e(W) = \gamma - X'(\beta^1 - \beta^0)$. In this case,

$$\Omega_e = \begin{bmatrix} E[g_1(W)g_1(W)'] & 0 & 0 & 0 \\ 0 & E[g_0(W)g_0(W)'] & 0 & 0 \\ 0 & 0 & E[m(W)m(W)'] & 0 \\ 0 & 0 & 0 & E[g_e(W)^2] \end{bmatrix}$$

To get the last row and column, we recall that $E[g_1(W)|X, Z, D] = E[g_0(W)|X, Z, D] = 0$ and that $E[m(W)|Z] = 0$. and $G_e = \begin{bmatrix} G & B \\ C & H \end{bmatrix}$ where G is as before, $B = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $C = \begin{bmatrix} \begin{pmatrix} -EX' & 0 \end{pmatrix} & \begin{pmatrix} EX' & 0 \end{pmatrix} & 0 \end{bmatrix}$, and $H = [1]$. Then, $G_e^{-1} = \begin{bmatrix} G^{-1} & 0 \\ -CG^{-1} & 1 \end{bmatrix}$ and asymptotic variance of $\widehat{ATE}$ equals
$E[g_e(W)^2] + CG^{-1}\Omega(G^{-1})'C' = E[X'(\beta^1 - \beta^0)(\beta^1 - \beta^0)'X] - \gamma^2 + CVC'$
$= Var(X'(\beta^1 - \beta^0)) + E(X')(V_{11} - V_{13} - V_{31} + V_{33})E(X)$.

3.2 Asymptotic Distribution of $\widehat{TT}$:

$$g_e(W) = D\left(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\frac{\phi(Z'\theta)}{\Phi(Z'\theta)}\right)$$

$$E\left[g_e(W)g_1(W)'|X,Z,D\right] = E\left[D\left(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\frac{\phi(Z'\theta)}{\Phi(Z'\theta)}\right)g_1(W)'|X,Z,D\right] = 0$$

because $E\left[g_1(W)'|X,Z,D\right]=0$. Then, $E(g_e(W)g_1(W)')=0$ by the Law of Iterated Expectations. On the other hand, by the same arguments, $E\left[g_0(W)'|X,Z,D\right]=0$ as well.
$E[g_e(W)m(W)'] = E\big[D\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)m(W)'\big]$
$= E\big[D\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)Z'\lambda_1(Z'\theta)\big]$ since $D(1-D)=0$, and $D^2=D$.
$= E_{X,Z}\big[\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)Z'\lambda_1(Z'\theta)E(D|X,Z)\big]$
$= E\big[\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)Z'\phi(Z'\theta)\big]$ since $E[D|Z]=\Phi(Z'\theta)$. So,

$$\Omega_e = \begin{bmatrix} E[g_1(W)g_1(W)'] & 0 & 0 & 0 \\ 0 & E[g_0(W)g_0(W)'] & 0 & 0 \\ 0 & 0 & E[m(W)m(W)'] & E[m(W)g_e(W)'] \\ 0 & 0 & E[g_e(W)m(W)'] & E[g_e(W)^2] \end{bmatrix}$$

Again, we write $G_e = \begin{bmatrix} G & B \\ C & H \end{bmatrix}$ where G, and B are as before, but $H = E[\Phi(Z'\theta)]$, and
$C = \begin{bmatrix} \begin{pmatrix} -E(X'\Phi(Z'\theta)) & -E[\phi(Z'\theta)] \end{pmatrix} & \begin{pmatrix} E(X'\Phi(Z'\theta)) & -E[\phi(Z'\theta)] \end{pmatrix} & -\tau \end{bmatrix} = \begin{bmatrix} C_1 & C_0 & -\tau' \end{bmatrix}$
with $\tau = E[Z(\alpha^1+\alpha^0)\big(\phi'(Z'\theta) - \phi(Z'\theta)\lambda_1(Z'\theta)\big)]$.
$E[g_e(W)^2] = E[D\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)^2] = P(D=1)E\big[\big(\gamma - X'(\beta^1-\beta^0) - (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)^2|D=1\big] = E[\Phi(Z'\theta)]\big(\gamma^2 - 2\gamma E[X'(\beta^1-\beta^0) + (\alpha^1+\alpha^0)\lambda_1(Z'\theta)|D=1] + E[\big(X'(\beta^1-\beta^0) + (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)^2|D=1]\big) = E[\Phi(Z'\theta)]\big(E[\big(X'(\beta^1-\beta^0) + (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)^2|D=1] - \gamma^2\big)$
The asymptotic variance of $\widehat{TT}$ is given by
$CVC' - 2\frac{1}{E[\Phi(Z'\theta)]}CG^{-1}T + \frac{1}{E[\Phi(Z'\theta)]}\left(E[\big(X'(\beta^1-\beta^0) + (\alpha^1+\alpha^0)\lambda_1(Z'\theta)\big)^2|D=1] - \gamma^2\right)$
where $T = \begin{pmatrix} 0 & 0 & E[g_e(W)m(W)'] \end{pmatrix}'$.

$$\begin{aligned} CVC' &= E(X'\Phi(Z'\theta))\big(V_{11} - 2V_{13} + V_{33}\big)E(X\Phi(Z'\theta)) \\ &+ 2E(X'\Phi(Z'\theta))\big(V_{12} + V_{14} - V_{32} - V_{34}\big)E(\phi(Z'\theta)) \\ &+ \big(E(\phi(Z'\theta))\big)^2\big(V_{22} + 2V_{24} + V_{44}\big) \\ &+ 2E(X'\Phi(Z'\theta))\big(V_{15} - V_{35}\big)\tau + 2E(\phi(Z'\theta))\big(V_{25} + V_{45}\big)\tau \\ &+ \tau' V_{55}\tau \end{aligned}$$

$$CG^{-1}T = -\big(C_1 G^{-1}_{1,\tilde\beta^1}G_{1,\theta} + C_0 G^{-1}_{0,\tilde\beta^0}G_{0,\theta} + \tau'\big)E\big(\Psi(W)g_e(W)'\big)$$

3.3 Asymptotic Distribution of $\widehat{LATE}$:

$$g_e(W) = \gamma - X(\beta^1 - \beta^0) - (\alpha^1 + \alpha^0)\frac{\phi(\tilde{z}'\theta) - \phi(z'\theta)}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)}$$

Let, $h(\tilde{z}'\theta, z'\theta) = \frac{\Phi(z'\theta)}{\Phi(\tilde{z}'\theta)}$. Then, $\frac{\mathrm{d}h}{\mathrm{d}\theta} = \big(z'\lambda_1(z'\theta) - \tilde{z}'\lambda_1(\tilde{z}'\theta)\big)h(\tilde{z}'\theta, z'\theta)$, and

$$g_e(W) = \gamma - X(\beta^1 - \beta^0) - (\alpha^1 + \alpha^0)\frac{\lambda_1(\tilde{z}'\theta) - \lambda_1(z'\theta)h(\tilde{z}'\theta, z'\theta)}{1 - h(\tilde{z}'\theta, z'\theta)}$$

$E\left[g_e(W)g_1(W)'|X,Z,D\right] = E\left[g_e(W)g_1(W)'|X,Z,D\right] = 0$ because $E\left[g_1(W)'|X,Z,D\right] = 0$. Then, $E(g_e(W)g_1(W)') = E_{X,Z,D}\Big[E\Big(g_e(W)g_1(W)'|X,Z,D\Big)\Big] = 0$ by the Law of Iterated Expectations. Similar arguments yield, $E[g_e(W)g_0(W)'] = 0$. Moreover, for this case, $E\left[g_e(W)m(W)'\right] = E_{X,Z}\left[g_e(W)E(m(W)'|X,Z)\right] = 0$ since now $E[m(W)|X,Z] = 0$. So Ω_e is again a block diagonal matrix.

On the other hand, G_e has the same form as before with $H = [1]$ and $C = \left[\begin{array}{ccc} C_1 & C_0 & -\tau' \end{array}\right]$ where

$C_1 = \left[\begin{array}{cc} -EX' & -r(\tilde{z}'\theta, z'\theta) \end{array}\right]$, $C_0 = \left[\begin{array}{cc} EX' & -r(\tilde{z}'\theta, z'\theta) \end{array}\right]$, $r(\tilde{z}'\theta, z'\theta) = \frac{\phi(\tilde{z}'\theta) - \phi(z'\theta)}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)}$ and

$\tau = (\alpha^1 + \alpha^0)\left(\frac{(\tilde{z}\phi'(\tilde{z}'\theta) - z\phi'(z'\theta))}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)} - r(\tilde{z}'\theta, z'\theta)\frac{(\tilde{z}\phi(\tilde{z}'\theta) - z\phi(z'\theta))}{\Phi(\tilde{z}'\theta) - \Phi(z'\theta)}\right)$

Asymptotic variance of $\widehat{LATE}$ equals $E[g_e(W)^2] + CG^{-1}\Omega(G^{-1})'C' = Var(X'(\beta^1 - \beta^0)) + CVC'$.

$$\begin{aligned}
CVC' &= E(X')\big(V_{11} - 2V_{13} + V_{33}\big)E(X) \\
&+ 2E(X')\big(V_{12} + V_{14} - V_{32} - V_{34}\big)r(\tilde{z}'\theta, z'\theta) \\
&+ \big(r(\tilde{z}'\theta, z'\theta)\big)^2\big(V_{22} + 2V_{24} + V_{44}\big) \\
&+ 2E(X')\big(V_{15} - V_{35}\big)\tau' + 2r(\tilde{z}'\theta, z'\theta)\big(V_{25} + V_{45}\big)\tau' \\
&+ \tau V_{55}\tau'
\end{aligned}$$

3.4 Asymptotic Distribution of $\widehat{MTE}$:

For MTE when $U^D = t$, $\hat{\gamma} := \frac{1}{n}\sum_{i=1}^{n} x_i'(\hat{\beta}^1 - \hat{\beta}^0) + (\hat{\alpha}_1 + \alpha_0)t$

$$g_e(W) = \gamma - X'(\beta^1 - \beta^0) - (\alpha^1 + \alpha^0)t$$

$E\left[g_e(W)g_1(W)'|X,Z,D\right] = \left(\gamma - X'(\beta^1 - \beta^0) - t\right)E[g_1(W)'|X,Z,D] = 0$. Then, by the law of Iterated Expectations, $E[g_e(W)g_1(W)'] = 0$. And symmetric arguments give us that $E[g_e(W)g_0(W)'] = 0$. At the same time, we have

$E\left[g_e(W)m(W)'\right] = E_{X,Z}\left[\left(\gamma - X(\beta^1 - \beta^0) - (\alpha^1 + \alpha^0)t\right)E(m(W)'|X,Z)\right] = 0$

Again Ω_e is a block diagonal matrix, and G_e has the same form as before with $C =$

$\left[\begin{array}{ccccc} -EX' & -t & EX' & -t & 0 \end{array}\right]$, $H = [1]$ and B as before.
Asymptotic variance of $\widehat{MTE}$ equals $E[g_e(W)^2] + CG^{-1}\Omega(G^{-1})'C' = Var(X'(\beta^1 - \beta^0)) + CVC'$.

$$\begin{aligned} CVC &= E(X')\big(V_{11} - 2V_{13} + V_{33}\big)E(X) \\ &+ 2E(X')\big(V_{12} + V_{14} - V_{32} - V_{34}\big)t \\ &+ t^2\big(V_{22} + 2V_{24} + V_{44}\big) \end{aligned}$$