Technical Appendix: "A GMM approach for dealing with missing data on regressors and instruments"

Jason Abrevaya
Department of Economics
University of Texas

Stephen G. Donald
Department of Economics
University of Texas

This version: April 2013

Proof of Lemma 2: Note that based on the first two moments, GMM solves

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(1-m_i)W_i(Y_i-W_i'\hat{\theta}) &= 0\\ \frac{1}{n}\sum_{i=1}^{n}(1-m_i)Z_i(X_i-Z_i'\hat{\gamma}) &= 0\end{aligned}$$

so the solutions are $\hat{\theta}_C$ and

$$\hat{\gamma}_C = \left(\sum_{i=1}^{n}(1-m_i)Z_iZ_i'\right)^{-1}\sum_{i=1}^{n}(1-m_i)Z_iX_i.$$

Based on the first and third conditions, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(1-m_i)W_i(Y_i-W_i'\hat{\theta}) &= 0\\ \frac{1}{n}\sum_{i=1}^{n}m_iZ_i(Y_i-Z_i'(\hat{\gamma}\hat{\alpha}+\hat{\beta})) &= 0\end{aligned}$$

so that again the solution for θ is $\hat{\theta}_C$, and then

$$\hat{\gamma} = \frac{1}{\hat{\alpha}_C}\left[\left(\sum_{i=1}^{n}(1-m_i)Z_iZ_i'\right)^{-1}\sum_{i=1}^{n}(1-m_i)Z_iY_i - \hat{\beta}_C\right]$$

when $\hat{\alpha}_C \neq 0$.

Proof of Proposition 2: By standard arguments the asymptotic variance covariance matrix of the GMM estimator of $\begin{pmatrix}\alpha_0 & \beta_0' & \gamma_0'\end{pmatrix}$ is given by $\left(G'\Omega^{-1}G\right)^{-1}$. Note that

$$\begin{aligned}G'\Omega^{-1}G &= \begin{pmatrix}G_{11} & 0\\ 0 & G_{22}\end{pmatrix}\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{12}' & \Omega_{22}\end{pmatrix}^{-1}\begin{pmatrix}G_{11} & 0\\ 0 & G_{22}\end{pmatrix}\\ &\quad+\begin{pmatrix}G_{31} & G_{32}\end{pmatrix}'\Omega_{33}^{-1}\begin{pmatrix}G_{31} & G_{32}\end{pmatrix}\end{aligned}$$

Then we can write

$$\begin{aligned}G'\Omega^{-1}G &= \Phi+\Upsilon\Delta\Upsilon'\\ \Phi &= \begin{pmatrix}G_{11} & 0\\ 0 & G_{22}\end{pmatrix}\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{12}' & \Omega_{22}\end{pmatrix}^{-1}\begin{pmatrix}G_{11} & 0\\ 0 & G_{22}\end{pmatrix}\\ \Upsilon' &= \begin{pmatrix}\gamma_0 & I & \alpha_0 I\end{pmatrix}\\ \Delta &= (1-\lambda)\Gamma_m\Omega_{\eta\eta m}^{-1}\Gamma_m\end{aligned}$$

using the facts that

$$\begin{aligned} \begin{pmatrix} G_{31} & G_{32} \end{pmatrix} &= -(1-\lambda)\Gamma_m \begin{pmatrix} \gamma_0 & I & \alpha_0 I \end{pmatrix} \\ \Omega_{33} &= (1-\lambda)\Omega_{\eta\eta m} \end{aligned}$$

The Woodbury matrix identity implies that

$$\left(G'\Omega^{-1}G\right)^{-1} = (\Phi + \Upsilon\Delta\Upsilon')^{-1} = \Phi^{-1} - \Phi^{-1}\Upsilon(\Delta^{-1} + \Upsilon'\Phi^{-1}\Upsilon)^{-1}\Upsilon'\Phi^{-1}$$

The asymptotic variance of GMM is the upper left $(1+K)\times(1+K)$ block of this matrix which corresponds to the asymptotic variance of the subvector $\hat{\theta} = (\hat{\alpha}, \hat{\beta}')$. Note that the first term on the right corresponding to Φ^{-1} is

$$\begin{pmatrix} G_{11}^{-1} & 0 \\ 0 & G_{22}^{-1} \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix} \begin{pmatrix} G_{11}^{-1} & 0 \\ 0 & G_{22}^{-1} \end{pmatrix} = \begin{pmatrix} G_{11}^{-1}\Omega_{11}G_{11}^{-1} & G_{11}^{-1}\Omega_{12}G_{22}^{-1} \\ G_{22}^{-1}\Omega_{12}'G_{11}^{-1} & G_{22}^{-1}\Omega_{22}G_{22}^{-1} \end{pmatrix}$$

and that the upper left $(1+K)\times(1+K)$ block of this corresponds to

$$G_{11}^{-1}\Omega_{11}G_{11}^{-1} = AVAR\left(\sqrt{n}(\hat{\theta}_C - \theta_0)\right)$$

so then

$$AVAR\left(\sqrt{n}(\hat{\theta}_C - \theta_0)\right) - AVAR(\sqrt{n}(\hat{\theta} - \theta_0))$$

equals the $(1+K)\times(1+K)$ upper left block of

$$\Phi^{-1}\Upsilon(\Delta^{-1} + \Upsilon'\Phi^{-1}\Upsilon)^{-1}\Upsilon'\Phi^{-1} \tag{0.1}$$

To find this, note that

$$\begin{aligned} G_{11} &= -E((1-m_i)W_iW_i') \\ &= -\lambda \begin{pmatrix} \gamma_0'\Gamma_c\gamma_0 + \sigma_{\xi c}^2 & \gamma_0'\Gamma_c \\ \Gamma_c\gamma_0 & \Gamma_c \end{pmatrix} \stackrel{def}{=} -\lambda\Gamma_c^W \end{aligned}$$

using (2.3) and Assumption 1, which gives

$$\begin{aligned} E(X_iZ_i'(1-m_i)) &= \lambda\gamma_0'\Gamma_c \\ E(X_i^2(1-m_i)) &= \lambda\left(\gamma_0'\Gamma_c\gamma_0 + \sigma_{\xi c}^2\right) \end{aligned}$$

so then

$$G_{11}^{-1} = -\frac{1}{\lambda}\left(\Gamma_c^W\right)^{-1} = -\frac{1}{\lambda}\left(\begin{array}{cc} \left(\sigma_{\xi c}^2\right)^{-1} & -\left(\sigma_{\xi c}^2\right)^{-1}\gamma_0' \\ -\left(\sigma_{\xi c}^2\right)^{-1}\gamma_0 & \Gamma_c^{-1} + \gamma_0\left(\sigma_{\xi c}^2\right)^{-1}\gamma_0' \end{array}\right)$$

Also,

$$G_{22}^{-1} = -\frac{1}{\lambda}\Gamma_c^{-1}$$

Then

$$\left(\begin{array}{cc} G_{11}^{-1} & 0 \\ 0 & G_{22}^{-1} \end{array}\right)\Upsilon = -\frac{1}{\lambda}\left(\begin{array}{c} \left(\begin{array}{c} 0 \\ \Gamma_c^{-1} \end{array}\right) \\ \alpha_0\Gamma_c^{-1} \end{array}\right)$$

so that

$$\Upsilon'\Phi^{-1}\Upsilon = \frac{1}{\lambda}\Gamma_c^{-1}\Omega_{\eta\eta c}\Gamma_c^{-1}$$

using

$$\begin{aligned} \Omega_{11} &= E(\varepsilon_i^2 W_i W_i'(1-m_i)) = \lambda\left(\begin{array}{cc} \gamma_0'\Omega_{\varepsilon\varepsilon c}\gamma_0 + 2\gamma_0'\Lambda_{\varepsilon\varepsilon\xi c} + \omega_{\varepsilon\xi c} & \gamma_0'\Omega_{\varepsilon\varepsilon c} + \Lambda_{\varepsilon\varepsilon\xi c}' \\ \Omega_{\varepsilon\varepsilon c}\gamma_0 + \Lambda_{\varepsilon\varepsilon\xi c} & \Omega_{\varepsilon\varepsilon c} \end{array}\right) \\ \Omega_{22} &= E(\varepsilon_i^2 Z_i Z_i'(1-m_i)) = \lambda\Omega_{\varepsilon\varepsilon c} \\ \Omega_{12}' &= E(\varepsilon_i\xi_i Z_i W_i'(1-m_i)) = \lambda\left(\begin{array}{cc} \Omega_{\varepsilon\xi c}\gamma_0 + \Lambda_{\varepsilon\xi\xi c} & \Omega_{\varepsilon\xi c} \end{array}\right) \end{aligned}$$

by (2.3) and Assumption 1. Also used is the fact that

$$\begin{aligned} \Lambda_{\varepsilon\eta\xi c} &= \Lambda_{\varepsilon\varepsilon\xi c} + \alpha_0\Lambda_{\varepsilon\xi\xi c} \\ \Omega_{\varepsilon\eta c} &= \Omega_{\varepsilon\varepsilon c} + \alpha_0\Omega_{\varepsilon\xi c} \\ \Omega_{\eta\eta c} &= \Omega_{\varepsilon\varepsilon c} + 2\alpha_0\Omega_{\varepsilon\xi c} + \alpha_0^2\Omega_{\xi\xi c} \end{aligned}$$

where each term (apart from α_0) is defined analogously to $\Lambda_{\varepsilon\eta\xi c}$ and $\Omega_{\varepsilon\eta c}$ in the text. Thus, we have

$$\Delta^{-1} + \Upsilon'\Phi^{-1}\Upsilon = \frac{1}{(1-\lambda)}\Gamma_m^{-1}\Omega_{\eta\eta m}\Gamma_m^{-1} + \frac{1}{\lambda}\Gamma_c^{-1}\Omega_{\eta\eta c}\Gamma_c^{-1}$$

Similarly,

$$\Upsilon'\Phi^{-1} = \frac{1}{\lambda}\left(\begin{array}{cc} \Gamma_c^{-1}\Omega_{\varepsilon\varepsilon c}^{Z,W}\left(\Gamma_c^W\right)^{-1} + \alpha_0\Gamma_c^{-1}\Omega_{\varepsilon,\xi c}^{Z,W}\left(\Gamma_c^W\right)^{-1} & \Gamma_c^{-1}\Omega_{\varepsilon,\xi c}\Gamma_c^{-1} + \Gamma_c^{-1}\Omega_{\xi\xi c}\Gamma_c^{-1} \end{array}\right)$$

Then the appropriate block of (0.1) is

$$\left(\Gamma_c^W\right)^{-1}\left(\Omega_{\varepsilon\varepsilon c}^{Z,W} + \alpha\Omega_{\varepsilon,\xi c}^{Z,W}\right)' D\left(\Omega_{\varepsilon\varepsilon c}^{Z,W} + \alpha\Omega_{\varepsilon,\xi c}^{Z,W}\right)\left(\Gamma_c^W\right)^{-1}$$

The result then follows from

$$\left(\Omega^{Z,W}_{\varepsilon\varepsilon c} + \alpha_0 \Omega^{Z,W}_{\varepsilon,\xi c}\right)\left(\Gamma^{W}_{c}\right)^{-1} = \left(\begin{array}{cc} \Gamma_c^{-1}\Lambda_{\varepsilon\eta\xi c}\left(\sigma^2_{\xi c}\right)^{-1} & \Gamma_c^{-1}\Omega_{\varepsilon\eta c}\Gamma_c^{-1} - \Gamma_c^{-1}\Lambda_{\varepsilon\eta\xi c}\left(\sigma^2_{\xi c}\right)^{-1}\gamma_0' \end{array}\right).$$

Proof of Lemma 4: Under Assumption 2, $A = 0$ and also $\Omega_{\varepsilon\eta c} = \sigma^2_{\varepsilon c}\Gamma_c$ so that $B = \sigma^2_{\varepsilon c} I$ from which the result follows using the general expression in Proposition 2.

Proof of Lemma 5: Under the conditions stated, $\Lambda_{\varepsilon\eta\xi c} = 0$ so that $A = 0$ and $B = \Gamma_c^{-1}\Omega_{\varepsilon\eta c}\Gamma_c^{-1}$ and the results follows.

Derivation of (2.12)-(2.15): Under the conditions in Assumption 4,

$$G_{11}^{-1} = -\frac{1}{\lambda}\left(\begin{array}{cc} \gamma_0'\Gamma\gamma_0 + \sigma^2_\xi & \gamma_0'\Gamma \\ \Gamma\gamma_0 & \Gamma \end{array}\right)^{-1}$$

and

$$\Omega_{11} = \lambda\sigma^2_\varepsilon \left(\begin{array}{cc} \gamma_0'\Gamma\gamma_0 + \sigma^2_\xi & \gamma_0'\Gamma \\ \Gamma\gamma_0 & \Gamma \end{array}\right)$$

so that

$$G_{11}^{-1}\Omega_{11}G_{11}^{-1} = \frac{\sigma^2_\varepsilon}{\lambda}\left(\begin{array}{cc} \left(\sigma^2_\xi\right)^{-1} & -\left(\sigma^2_\xi\right)^{-1}\gamma_0' \\ -\left(\sigma^2_\xi\right)^{-1}\gamma_0 & \Gamma^{-1} + \gamma_0\left(\sigma^2_\xi\right)^{-1}\gamma_0' \end{array}\right) \tag{0.2}$$

which gives (2.12) and (2.13). For (2.14) and (2.15) under Assumption 4, we have $A = 0$ and $B = \sigma^2_\varepsilon I$ while

$$D = \frac{1}{\lambda^2}\left((\sigma^2_\varepsilon + \alpha_0^2\sigma^2_\xi)\left(\frac{1}{(1-\lambda)} + \frac{1}{\lambda}\right)\right)^{-1}\Gamma^{-1}$$

so that

$$\left(\begin{array}{c} A' \\ B' \end{array}\right) D \left(\begin{array}{cc} A & B \end{array}\right) = \frac{1}{\lambda}\frac{\sigma^4_\varepsilon(1-\lambda)}{(\sigma^2_\varepsilon + \alpha_0^2\sigma^2_\xi)}\left(\begin{array}{cc} 0 & 0 \\ 0 & \Gamma^{-1} \end{array}\right)$$

from which (2.14) and (2.15) follow using Proposition 2 and (0.2).

Proof of Proposition 3: It is convenient to stack the observations according to whether they are complete or not so that the GD estimator can then be thought of as GLS for the regression

$$\begin{aligned} \left(\begin{array}{c} Y_c \\ Y_m \end{array}\right) &= \left(\begin{array}{c} X_c \\ \hat{X}_m \end{array}\right)\alpha_0 + \left(\begin{array}{c} Z_c \\ Z_m \end{array}\right)\beta_0 + \left(\begin{array}{c} \varepsilon_c \\ \varepsilon_m + \alpha_0\xi_m + \alpha_0 Z_m(\gamma_0 - \hat{\gamma}) \end{array}\right) \\ &= \left(\begin{array}{c} W_c \\ \hat{W}_m \end{array}\right)\theta_0 + \left(\begin{array}{c} \varepsilon_c \\ \eta_m + \alpha_0 Z_m(\gamma_0 - \hat{\gamma}) \end{array}\right) \\ \hat{X}_m &= Z_m\hat{\gamma},\ \hat{\gamma} = (Z_c'Z_c)^{-1}Z_c'X_c \end{aligned}$$

using the inverse of the estimated variance covariance matrix of the residuals

$$\Psi = \begin{pmatrix} \sigma_\varepsilon^2 I & 0 \\ 0 & \left(\sigma_\varepsilon^2 + \alpha^2\sigma_\xi^2\right) I + \sigma_\xi^2\alpha^2 Z_m(Z_c'Z_c)^{-1}Z_m' \end{pmatrix}$$

Use $\sigma_\varepsilon^2, \alpha_0, \sigma_\xi^2$ as the limiting values of the estimates used in this variance covariance matrix, and let $\sigma_\eta^2 = \sigma_\varepsilon^2 + \alpha_0^2\sigma_\xi^2$. Then one can see that $\sqrt{n}(\hat{\theta}_{GD} - \theta_0)$ is asymptotically equivalent to

$$\left[\frac{1}{n}\begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}' \Psi^{-1} \begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}\right]^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}' \Psi^{-1} \begin{pmatrix} \varepsilon_c \\ \varepsilon_m + \alpha_0\xi_m + \alpha_0 Z_m(\gamma - \hat{\gamma}) \end{pmatrix} \tag{0.3}$$

where $\bar{W}_m$ is just $\hat{W}_m$ with $\hat{\gamma}$ replaced with its probability limit γ_0. Note that

$$\begin{aligned} \sigma_\eta^2 I + \sigma_\xi^2\alpha_0^2 Z_m(Z_c'Z_c)^{-1}Z_m' &= \frac{1}{\sigma_\eta^2}(I - Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m') \\ b &= \frac{\sigma_\eta^2}{\sigma_\xi^2\alpha_0^2} \end{aligned}$$

so then

$$\frac{1}{n}\begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}' \Psi^{-1} \begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix} \tag{0.4}$$

$$= \frac{1}{n}\left(\frac{1}{\sigma_\varepsilon^2}W_c'W_c + \frac{1}{\sigma_\eta^2}\bar{W}_m\bar{W}_m - \frac{1}{\sigma_\eta^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m'\bar{W}_m\right) \overset{def}{=} H_{GD} \tag{0.5}$$

and

$$\begin{aligned} &\begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}' \Psi^{-1} \begin{pmatrix} \varepsilon_c \\ \eta_m + \alpha_0 Z_m(\gamma_0 - \hat{\gamma}) \end{pmatrix} \\ = &\frac{1}{\sigma_\varepsilon^2}W_c'\varepsilon_c + \frac{1}{\sigma_\eta^2}\bar{W}_m\eta_m - \alpha_0\frac{1}{\sigma_\eta^2}\bar{W}_m Z_m\left(Z_c'Z_c\right)^{-1}Z_c'\varepsilon_c \\ &+\alpha_0\frac{1}{\sigma_\eta^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m'Z_m\left(Z_c'Z_c\right)^{-1}Z_c'\varepsilon_c \\ &-\frac{1}{\sigma_\eta^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m'\eta_m \end{aligned}$$

using $\gamma_0 - \hat{\gamma} = -(Z_c'Z_c)^{-1}Z_c'\varepsilon_c$. But,

$$\begin{aligned} &\alpha_0\frac{1}{\sigma_\eta^2}\bar{W}_m Z_m\left(Z_c'Z_c\right)^{-1}Z_c'\varepsilon_c \\ &-\alpha_0\frac{1}{\sigma_\eta^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m'Z_m\left(Z_c'Z_c\right)^{-1}Z_c'\varepsilon_c \\ = &\frac{1}{\alpha_0\sigma_\xi^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_c'\varepsilon_c \end{aligned} \tag{0.6}$$

so that

$$\frac{1}{\sqrt{n}}\begin{pmatrix} W_c \\ \bar{W}_m \end{pmatrix}' \Psi^{-1}\begin{pmatrix} \varepsilon_c \\ \eta_m + \alpha_0 Z_m(\gamma - \hat{\gamma}) \end{pmatrix} \tag{0.7}$$

$$= \frac{1}{\sqrt{n}}(\frac{1}{\sigma_\varepsilon^2}W_c'\varepsilon_c + \frac{1}{\sigma_\eta^2}\bar{W}_m\eta_m$$

$$- \frac{1}{\alpha_0\sigma_\xi^2}\bar{W}_m Z_m(\frac{1}{b}Z_c'Z_c + Z_m'Z_m)^{-1}Z_c'\varepsilon_c - \frac{1}{\sigma_\eta^2}\bar{W}_m Z_m(bZ_c'Z_c + Z_m'Z_m)^{-1}Z_m'\eta_m$$

For GMM, using similar notation, one can show that $\sqrt{n}(\hat{\theta} - \theta_0)$ behaves asymptotically like the appropriate elements of the following vector,

$$\left(\frac{1}{n}\begin{pmatrix} W_c'W_c & 0 \\ 0 & Z_c'Z_c \\ Z_m'\bar{W}_m & Z_m'Z_m\alpha_0 \end{pmatrix}'\begin{pmatrix} \frac{1}{\sigma_\varepsilon^2}(W_c'W_c)^{-1} & 0 & 0 \\ 0 & \frac{1}{\sigma_\xi^2}(Z_c'Z_c)^{-1} & 0 \\ 0 & 0 & \frac{1}{\sigma_\eta^2}(Z_m'Z_m)^{-1} \end{pmatrix}\begin{pmatrix} W_c'W_c & 0 \\ 0 & Z_c'Z_c \\ Z_m'\bar{W}_m & Z_m'Z_m\alpha_0 \end{pmatrix}\right.$$

$$\times \begin{pmatrix} W_c'W_c & 0 \\ 0 & Z_c'Z_c \\ Z_m'\bar{W}_m & Z_m'Z_m\alpha_0 \end{pmatrix}'\begin{pmatrix} \frac{1}{\sigma_\varepsilon^2}(W_c'W_c)^{-1} & 0 & 0 \\ 0 & \frac{1}{\sigma_\xi^2}(Z_c'Z_c)^{-1} & 0 \\ 0 & 0 & \frac{1}{\sigma_\eta^2}(Z_m'Z_m)^{-1} \end{pmatrix}\frac{1}{\sqrt{n}}\begin{pmatrix} W_c'\varepsilon_c \\ Z_c'\varepsilon_c \\ Z_m'\eta_m \end{pmatrix}$$

Note that the second term corresponds to the score and can be shown to equal

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{\sigma_\varepsilon^2}W_c'\varepsilon_c + \frac{1}{\sigma_\eta^2}\hat{W}_m'\eta_m \\ \frac{1}{\sigma_\xi^2}Z_c'\varepsilon_c + \alpha_0 Z_m'\eta_m \end{pmatrix} \tag{0.8}$$

From the first term, which corresponds to the Hessian, we require the first $(1 + K)$ rows that correspond to θ_0 and these turn out to be equal to

$$\begin{pmatrix} (H_{GD})^{-1} & -(H_{GD})^{-1}\hat{W}_m'Z_m\frac{1}{\alpha_0}(bZ_c'Z_c + Z_m'Z_m)^{-1} \end{pmatrix} \tag{0.9}$$

The result then follows by noticing that (0.4) multiplied by (0.7) is identical to (0.9) multiplied by (0.8).

Proof of Proposition 4: The GMM estimator using $\hat{\Omega}_S^{-1}$ (after the usual standardization) behaves like the quantity

$$\left(\frac{1}{n}\begin{pmatrix} W_c'W_c & \bar{W}_m Z_m \end{pmatrix}\Omega_S^{-1}\begin{pmatrix} W_c'W_c \\ Z_m'\bar{W}_m \end{pmatrix}\right)^{-1} \times$$

$$\begin{pmatrix} W_c'W_c & \bar{W}_m Z_m \end{pmatrix}\Omega_S^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} W_c'\varepsilon_c \\ Z_m'\eta_m + \alpha_0 Z_m'Z_m(\gamma_0 - \hat{\gamma}) \end{pmatrix}$$

where we have replaced estimates with limits in Ω_S^{-1}. Using the fact that

$$\left(\sigma_\eta^2 Z_m' Z_m + \alpha_0^2\sigma_\xi^2 (Z_m' Z_m) \left(Z_c' Z_c\right)^{-1} Z_m' Z_m\right)^{-1} = \frac{1}{\sigma_\eta^2}\left(Z_m' Z_m\right)^{-1} - \frac{1}{\sigma_\eta^2}\left(b Z_c' Z_c + Z_m' Z_m\right)^{-1}$$

one can show that the first (Hessian) term equals H_{GD} while the second (score) term equals the expression in (0.7) using (0.6).

Proof of Proposition 6: Even under stronger assumptions that $E(\varepsilon_i|X_i, Z_i, m_i) = E(\xi_i|Z_i, m_i) = 0$, this estimator is biased due to the omitted variable $m_i Z_{2i}$ from equation (3.22) except in the case where $\gamma_{02}\alpha_0 = 0$ which requires either (i) or (ii) in the statement of the Proposition. To characterize the inconsistency, it is convenient to stack that data as in Propositions 3-5 so that we can write the Dummy Variable estimator of the entire vector $(\alpha_0, \beta_0', \delta_0)$ (where we use $\delta_0 = \gamma_{01}\alpha_0$ to represent the population coefficient on the included dummy m_i) in (3.22)),

$$\begin{pmatrix} \hat\alpha_{DM} \\ \hat\beta_{DM} \\ \hat\delta \end{pmatrix} = \left(\begin{pmatrix} X_c & Z_c & 0 \\ 0 & Z_m & \iota \end{pmatrix}' \begin{pmatrix} X_c & Z_c & 0 \\ 0 & Z_m & \iota \end{pmatrix}\right)^{-1} \times \begin{pmatrix} X_c & Z_c & 0 \\ 0 & Z_m & \iota \end{pmatrix}' \begin{pmatrix} Y_c \\ Y_m \end{pmatrix}$$

where ι is a column vector of ones – ie the last column in the rgressor matrix gives the vector corresponding to m_i stacked by the two groups. By the Frisch-Waugh-Lovell Theorem, we can write

$$\begin{pmatrix} \hat\alpha_{DM} \\ \hat\beta_{DM} \end{pmatrix} = \left(\begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}\right)^{-1} \times \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} Y_c \\ Y_m \end{pmatrix}$$

where

$$M_m = I - \frac{1}{n_m}\begin{pmatrix} 0 \\ \iota \end{pmatrix}\begin{pmatrix} 0 & \iota' \end{pmatrix}$$

Under the normalization on the second moment matrix for Z,

$$\Gamma = \begin{pmatrix} 1 & 0' \\ 0 & \Gamma_{22} \end{pmatrix} = \begin{pmatrix} e_1' \\ \Gamma_{2\bullet} \end{pmatrix} = \begin{pmatrix} e_1 & \Gamma_{2\bullet}' \end{pmatrix}$$

$$\Gamma_{2\bullet} = \begin{pmatrix} 0 & \Gamma_{22} \end{pmatrix},\ e_1' = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$$

Then,

$$\begin{pmatrix} \hat{\alpha}_{DM} \\ \hat{\beta}_{DM} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} + \left(\begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix} \right)^{-1} \times \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} \varepsilon_c \\ Z_{m,2}\gamma_{02}\alpha_0 + \varepsilon_m + \alpha_0 \xi_m \end{pmatrix}$$

One can show

$$\frac{1}{n} \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}$$

and

$$\begin{aligned} & \frac{1}{n} \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} \varepsilon_c \\ Z_{m,2}\gamma_{02}\alpha_0 + \varepsilon_m + \alpha_0 \xi_m \end{pmatrix} \\ = & \frac{1}{n} \begin{pmatrix} 0 \\ \left(Z_m' Z_{m,2} - \frac{\iota' Z_m}{n_m}\iota' Z_{m,2}\right)\gamma_{02}\alpha_0 \end{pmatrix} + o(1) \\ & \xrightarrow{p} \begin{pmatrix} 0 \\ \lambda\Gamma_{2\bullet}'\gamma_{02}\alpha_0 \end{pmatrix} \end{aligned}$$

so that

$$\begin{pmatrix} \hat{\alpha}_{DM} \\ \hat{\beta}_{DM} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \lambda\Gamma_{2\bullet}'\gamma_{02}\alpha_0 \end{pmatrix}$$

Thus, the estimator is inconsistent unless $\gamma_{02}\alpha_0 = 0$.

Proof of Proposition 7: (i) Using the arguments in the proof of Proposition 6, one can show that when $\alpha_0 = 0$,

$$\sqrt{n}(\hat{\theta}_{DM} - \theta) = \begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} \varepsilon_c \\ \varepsilon_m \end{pmatrix} + o_p(1)$$

Now,

$$\frac{1}{\sqrt{n}} \begin{pmatrix} X_c & Z_c \\ 0 & Z_m \end{pmatrix}' M_m \begin{pmatrix} \varepsilon_c \\ \varepsilon_m \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} X_c'\varepsilon_c \\ Z_c'\varepsilon_c + Z_m'(I - \frac{1}{n_m}\iota\iota')\varepsilon_m \end{pmatrix} \tag{0.10}$$

and the asymptotic variance of the right hand side is then

$$\sigma_\varepsilon^2 \begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}$$

so then the asymptotic variance of the DM estimator in case (i) is

$$\sigma_\varepsilon^2 \begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}^{-1}$$

The result in (i) follows using the fact that

$$(\Gamma - (1-\lambda)e_1e_1')^{-1} = \begin{pmatrix} \lambda & 0' \\ 0 & \Gamma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \lambda^{-1} & 0' \\ 0 & \Gamma_{22}^{-1} \end{pmatrix}$$

and the partitioned inverse formula, which gives

$$\begin{pmatrix} \lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \lambda\gamma_0'\Gamma \\ \lambda\Gamma\gamma_0 & \Gamma - (1-\lambda)e_1e_1' \end{pmatrix}^{-1} = \begin{pmatrix} d^{-1} & -d^{-1}\begin{pmatrix} \gamma_{01} \\ \lambda\gamma_{02} \end{pmatrix}' \\ -d^{-1}\begin{pmatrix} \gamma_{01} \\ \lambda\gamma_{02} \end{pmatrix} & D \end{pmatrix} \quad (0.11)$$

where

$$\begin{aligned} d &= \lambda\left(\sigma_\xi^2 + (1-\lambda)\gamma_{02}'\Gamma_{22}\gamma_{02}\right) \\ D &= \begin{pmatrix} \lambda^{-1} & 0' \\ 0 & \Gamma_{22}^{-1} \end{pmatrix} + d^{-1}\lambda^2 \begin{pmatrix} \lambda^{-2}\gamma_{01}^2 & \lambda^{-1}\gamma_{01}\gamma_{02}' \\ \lambda^{-1}\gamma_{02}\gamma_{01} & \gamma_{02}\gamma_{02}' \end{pmatrix} \end{aligned}$$

For (ii), the term corresponding to (0.10) is

$$\frac{1}{\sqrt{n}} \begin{pmatrix} X_c'\varepsilon_c \\ Z_c'\varepsilon_c + Z_m'(I - \frac{1}{n_m}\iota\iota')\eta_m \end{pmatrix}$$

which has asymptotic variance

$$\begin{pmatrix} \sigma_\varepsilon^2\lambda\left(\gamma_0'\Gamma\gamma_0 + \sigma_\xi^2\right) & \sigma_\varepsilon^2\lambda\gamma_0'\Gamma \\ \sigma_\varepsilon^2\lambda\Gamma\gamma_0 & \sigma_\varepsilon^2\lambda\Gamma + \sigma_\eta^2(1-\lambda)(\Gamma - e_1e_1') \end{pmatrix}$$

Imposing the restriction that $\gamma_{02} = 0$ and pre- and post-multiplying this by (0.11) gives the result.

Proofs for Proposition 8 and 9: For the complete-data IV estimator, the asymptotic variance is given by

$$\sigma_\varepsilon^2 \left(p\lim \frac{1}{n} \begin{pmatrix} Y_{2c}'P_{W_c}Y_{2c} & Y_{2c}'Z_c \\ Z_c'Y_{2c} & Z_c'Z_c \end{pmatrix}\right)^{-1} \quad (0.12)$$

where we use similar notation to the prevous results so that $W_c = \begin{pmatrix} X_c & Z_c \end{pmatrix}$. Under the normalization,

$$\begin{aligned} &\frac{1}{n}Z_c'Z_c \xrightarrow{p} \lambda I \\ &\frac{1}{n}Z_c'Y_{2c} \xrightarrow{p} \lambda\left(\gamma_0\pi_0 + \Pi_0\right) \\ &\frac{1}{n}Y_{2c}'P_{W_c}Y_{2c} \xrightarrow{p} \lambda \begin{pmatrix} \pi_0 \\ \Pi_0 \end{pmatrix}' \begin{pmatrix} \gamma_0'\gamma_0 + \sigma_\xi^2 & \gamma_0' \\ \gamma_0 & I \end{pmatrix} \begin{pmatrix} \pi_0 \\ \Pi_0 \end{pmatrix} \\ = \; &\lambda\pi_0'\left(\gamma_0'\gamma_0 + \sigma_\xi^2\right)\pi_0 + 2\lambda\Pi_0'\gamma_0\pi_0 + \lambda\Pi_0'\Pi_0 \end{aligned}$$

The results for $\sqrt{n}(\hat{\delta}_C - \delta)$ and $\sqrt{n}(\hat{\beta}_C - \beta)$ then follow from these results and (0.12) using the partitioned inverse formula. For the Full instrument estimator and the Dummy estimator, we do similar calculations based on asymptotic variances given by

$$\sigma_\varepsilon^2\left(p\lim\frac{1}{n}\begin{pmatrix} Y_2'P_WY_2 & Y_2'Z \\ Z'Y_2 & Z'Z \end{pmatrix}\right)^{-1} \tag{0.13}$$

for appropriate projection matrices P_W that differ for the two estimators. Note that since

$$Y_{2i} = (1-m_i)X_i\pi_0 + Z_i'\Pi_0 + m_i\gamma_{01}\pi_0 + m_iZ_{2i}'\gamma_{02}\pi_0 + v_i + \pi_0m_i\omega_i$$

we can stack the “complete” and “missing” data separately and write

$$\begin{pmatrix} Y_{2c} \\ Y_{2m} \end{pmatrix} = \begin{pmatrix} X_c & Z_c & 0 & 0 \\ 0 & Z_m & \iota & Z_{2m} \end{pmatrix}\begin{pmatrix} \pi_0 \\ \Pi_0 \\ \gamma_{01}\pi_0 \\ \gamma_{02}\pi_0 \end{pmatrix} + \begin{pmatrix} v_c \\ v_m + \pi_0\omega_m \end{pmatrix} \tag{0.14}$$

For both the Dummy and Full instrument estimators, we have

$$\frac{1}{n}Z'Y_2 = \frac{1}{n}Z_c'Y_{2c} + \frac{1}{n}Z_m'Y_{2m} \xrightarrow{p} \gamma_0\pi_0 + \Pi_0 \tag{0.15}$$

The term $\frac{1}{n}Y_2'P_WY_2$ differs for the Full and Dummy Instrument set estimators. To distinguish them, use the notation P_{W_F} for the projection using the Full Instrument set and P_{W_D} for the dummy instrument set. For the Full instrument set, noting that $Z_m = \begin{pmatrix} \iota & Z_{2m} \end{pmatrix}$ and $\gamma\pi_0 = \begin{pmatrix} \gamma_{01}\pi_0 & \gamma_{02}'\pi_0 \end{pmatrix}$, we would have

$$\begin{pmatrix} \pi_0 & \Pi_0' & \gamma_0'\pi_0 \end{pmatrix}\begin{pmatrix} \lambda\gamma_0'\gamma_0 + \lambda\sigma_\xi^2 & \lambda\gamma_0' & 0 \\ \lambda\gamma_0 & I & (1-\lambda)I \\ 0 & (1-\lambda)I & (1-\lambda)I \end{pmatrix}\begin{pmatrix} \pi_0 \\ \Pi_0 \\ \gamma_0\pi_0 \end{pmatrix}$$

$$\frac{1}{n}Y_2'P_{W_F}Y_2 \xrightarrow{p} \pi_0^2\lambda\sigma_\xi^2 + \left(\pi_0\gamma_0' + \Pi_0'\right)\left(\pi_0\gamma_0 + \Pi_0\right) \tag{0.16}$$

Then, the results for the asymptotic variances of $\sqrt{n}(\hat{\delta}_F - \delta_0)$ and $\sqrt{n}(\hat{\beta}_F - \beta_0)$ follow from (0.13), (0.15) and (0.16) using the partitioned inverse formula. For the Dummy instrument set, the only difference is that now in place of (0.16) we have

$$\frac{1}{n}Y_2'P_{W_D}Y_2 \xrightarrow{p} \pi_0^2\lambda\sigma_\xi^2\left[1 - \frac{(1-\lambda)\gamma_{02}'\gamma_{02}}{\sigma_\xi^2 + (1-\lambda)\gamma_{02}'\gamma_{02}}\right] + \left(\pi_0\gamma_0' + \Pi_0'\right)\left(\pi_0\gamma_0 + \Pi_0\right) \tag{0.17}$$

To derive this, we use the fact that

$$Y_2'P_{W_D}Y_2 = Y_2'P_{W_F}Y_2 - Y_2'M_{W_D}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}\left(\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'M_{W_D}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}\right)^{-1}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'M_{W_D}Y_2$$

where $M_{W_D} = I - P_{W_D}$. In doing the calculation on the second term on the right hand side, it is convenient to rewrite (0.14) as

$$\begin{pmatrix} Y_{2c} \\ Y_{2m} \end{pmatrix} = \begin{pmatrix} Z_c & X_c & 0 & 0 \\ Z_m & 0 & \iota & Z_{2m} \end{pmatrix}\begin{pmatrix} \Pi_0 \\ \pi_0 \\ \gamma_{01}\pi_0 \\ \gamma_{02}\pi_0 \end{pmatrix} + \begin{pmatrix} v_c \\ v_m + \pi_0\omega_m \end{pmatrix}$$

We can write the dummy instrument set as

$$W_D = \begin{pmatrix} Z_c & X_c & 0 \\ Z_m & 0 & \iota \end{pmatrix}$$

Then the result involves the fact that

$$\left(\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'M_{W_D}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}\right)^{-1}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'M_{W_D}Y_2 \xrightarrow{p} \gamma_{02}\pi_0$$

and that

$$\frac{1}{n}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'M_{W_D}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix} - \frac{1}{n}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'P_{W_D}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}$$

From here,

$$\frac{1}{n}\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix}'\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix} \xrightarrow{p} (1-\lambda)I_{K-1}$$

while

$$\frac{1}{n}W_D'\begin{pmatrix} 0 \\ Z_{2m} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} (1-\lambda)\begin{pmatrix} 0 \\ I_{K-1} \end{pmatrix} \\ 0 \\ 0 \end{pmatrix}$$

and

$$\frac{1}{n}W_D'W_D \xrightarrow{p} \begin{pmatrix} I_K & \lambda\gamma_0 & (1-\lambda)e_1 \\ \lambda\gamma_0' & \lambda\gamma_0'\gamma_0 + \lambda\sigma_\xi^2 & 0 \\ (1-\lambda)e_1' & 0 & (1-\lambda) \end{pmatrix}$$

Then, applying the partitioned inverse formula gives the expression in (0.17). Further application of the partitioned inverse formula gives the results for $\sqrt{n}(\hat{\delta}_D - \delta_0)$ and $\sqrt{n}(\hat{\beta}_D - \beta_0)$.