B Supplemental web appendix for "Semiparametrically efficient estimation of the average linear regression function" by Bryan Graham and Cristine Pinto

This supplemental web appendix contains proofs of the results not included in the main appendix as well as additional detailed calculations for some proof steps. All notation is as defined in the main text and/or appendix unless explicitly noted otherwise. Equation numbering continues in sequence with that established in the main text and its appendix.

Proof of Proposition 2

Begin by noting that under Assumption 4 we have

$$\begin{aligned} h(x,U) &= \underline{h}(U) + \int_{\underline{x}}^{x} \frac{\partial h(t,U)}{\partial x} \mathrm{d}t \\ &= \underline{h}(U) + \int_{\underline{x}}^{\bar{x}} \frac{\partial h(t,U)}{\partial x} \mathbf{1}(x \geq t)\, \mathrm{d}t, \end{aligned}$$

which, invoking conditional independence yields

$$\begin{aligned} \mathbb{E}\left[\frac{X - e_0(W)}{v_0(W)} \underline{h}(U)\right] &= \mathbb{E}\left[\frac{X - e_0(W)}{v_0(W)} \mathbb{E}\left[\left.\underline{h}(U)\right| W, X\right]\right] \\ &= \mathbb{E}\left[\frac{X - e_0(W)}{v_0(W)} \mathbb{E}\left[\left.\underline{h}(U)\right| W\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left.\frac{X - e_0(W)}{v_0(W)}\right| W\right] \mathbb{E}\left[\left.\underline{h}(U)\right| W\right]\right] \\ &= \mathbb{E}\left[v_0(W)^{-1} \mathbb{E}\left[\left.X - e_0(W)\right| W\right] \mathbb{E}\left[\left.\underline{h}(U)\right| W\right]\right] \\ &= \mathbb{E}\left[v_0(W)^{-1} \cdot 0 \cdot \mathbb{E}\left[\left.\underline{h}(U)\right| W\right]\right] \\ &= 0. \end{aligned}$$

Using this result we can re-write the β_0 estimand as follows:

$$
\begin{aligned}
\mathbb{E}\left[\frac{X-e_0(W)}{v_0(W)}Y\right] &= \mathbb{E}\left[\frac{X-e_0(W)}{v_0(W)}\int_{\underline{x}}^{\bar{x}}\frac{\partial h(t,U)}{\partial x}\mathbf{1}(X\geq t)\,\mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\underline{x}}^{\bar{x}}\frac{\partial h(t,U)}{\partial x}\mathbf{1}(X\geq t)\frac{X-e_0(W)}{v_0(W)}\mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[\left.\frac{\partial h(t,U)}{\partial x}\right|W,X\right]\mathbf{1}(X\geq t)\frac{X-e_0(W)}{v_0(W)}\mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[\left.\frac{\partial h(t,U)}{\partial x}\right|W\right]\mathbf{1}(X\geq t)\frac{X-e_0(W)}{v_0(W)}\mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[\left.\frac{\partial h(t,U)}{\partial x}\right|W\right]\mathbb{E}\left[\left.\frac{X-e_0(W)}{v_0(W)}\right|W,X\geq t\right]\left(1-F_{X|W}(t|W)\right)\mathrm{d}t\right].
\end{aligned}
$$

Next observe that

$$
\begin{aligned}
v_0(w) &= \mathbb{E}\left[X\left(X-e_0(W)\right)|W=w\right]\\
&= \mathbb{E}\left[\left.\int_{\underline{x}}^{\bar{x}}\left(X-e_0(W)\right)\mathrm{d}t\right|W=w\right]\\
&= \int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[X-e_0(W)|W=w,X\geq t\right]\left(1-F_{X|W}(t|w)\right)\mathrm{d}t.
\end{aligned}
$$

Putting all these pieces together we have

$$
\begin{aligned}
\mathbb{E}\left[\frac{X-e_0(W)}{v_0(W)}Y\right] =&\mathbb{E}\left[\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[\left.\frac{\partial h(t,U)}{\partial x}\right|W\right]\frac{\mathbb{E}\left[X-e_0(W)|W,X\geq t\right]\left(1-F_{X|W}(t|W)\right)}{\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[X-e_0(W)|W,X\geq v\right]\left(1-F_{X|W}(v|W)\right)\mathrm{d}v}\mathrm{d}t\right]\\
=&\mathbb{E}\Bigg[\int_{\underline{x}}^{\bar{x}}\int_{-\infty}^{\infty}\Bigg(\frac{\partial h(t,u)}{\partial x}f_{U|W,X}(u|w,t)\,\mathrm{d}u\\
&\times\frac{1}{f_{X|W}(t|W)}\frac{\mathbb{E}\left[X-e_0(W)|W,X\geq t\right]\left(1-F_{X|W}(t|W)\right)}{\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[X-e_0(W)|W,X\geq v\right]\left(1-F_{X|W}(v|W)\right)\mathrm{d}v}f_{X|W}(t|W)\Bigg)\,\mathrm{d}t\Bigg]\\
=&\mathbb{E}\Bigg[\int_{\underline{x}}^{\bar{x}}\int_{-\infty}^{\infty}\Bigg(\frac{\partial h(t,u)}{\partial x}\frac{1}{f_{X|W}(t|W)}\\
&\times\frac{\mathbb{E}\left[X-e_0(W)|W,X\geq t\right]\left(1-F_{X|W}(t|W)\right)}{\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[X-e_0(W)|W,X\geq v\right]\left(1-F_{X|W}(v|W)\right)\mathrm{d}v}f_{U,X|W}(u,t|W)\Bigg)\,\mathrm{d}u\mathrm{d}t\Bigg]\\
=& \mathbb{E}\left[\omega(W,X)\frac{\partial h(X,U)}{\partial x}\right],
\end{aligned}
$$

with

$$\omega(w,x)=\frac{1}{f_{X|W}(x|w)}\frac{\mathbb{E}\left[X-e_0(W)|W=w,X\geq x\right]\left(1-F_{X|W}(x|w)\right)}{\int_{\underline{x}}^{\bar{x}}\mathbb{E}\left[X-e_0(W)|W=w,X\geq v\right]\left(1-F_{X|W}(v|w)\right)\mathrm{d}v}.$$

Proof of Corollary 1

Let $f(x|w;\phi)$ be a known parametric family of conditional distributions for X given W. Let $f_0(x|w)=f(x|w;\phi)$ at some unique $\phi=\phi_0$. Relative to that considered in Theorem 1, the parametric submodel changes to

$$f(w,x,y;\eta)=f(y|w,x)\,f(x|w;\phi(\eta))\,f(w;\eta)$$

with an associated score vector of

$$s_\eta(w,x,y;\eta)=s_\eta(y|w,x;\eta)+\left(\frac{\partial\phi(\eta)}{\partial\eta'}\right)'\mathbb{S}_\phi(x|w;\phi)+t_\eta(w;\eta), \tag{56}$$

where $\mathbb{S}_\phi(x|w;\phi)$ is the score function associated with the parametric conditional log-likelihood for ϕ.

From (56), and the usual (conditional) mean zero properties of score functions, the tangent set is evidently

$$\mathcal{T}=\left\{s(y|w,x)+\mathbf{c}\mathbb{S}_\phi(x|w)+t(w)\right\}$$

where $\mathbb{S}_\phi(x|w)=\mathbb{S}_\phi(x|w;\phi_0)$, $\mathbf{c}$ is a matrix of constants, and

$$\mathbb{E}\left[s(Y|W,X)|W,X\right]=\mathbb{E}\left[\mathbb{S}_\phi(X|W)|W\right]=\mathbb{E}\left[t(W)\right]=0.$$

To show pathwise differentiability, begin by noting that $\beta(\eta)$ continues to equal (37), however $b(w;\eta)$ now satisfies the modified conditional moment restriction

$$\int\int\begin{pmatrix}1\\x\end{pmatrix}(y-a(w;\eta)-x'b(w;\eta))\,f(y|w,x;\eta)\,f(x|w;\phi(\eta))\,\mathrm{d}x\mathrm{d}y=0. \tag{57}$$

We can derive a close-form expression for $\frac{\partial b(w;\eta_0)}{\partial\eta'}$ in (39) by differentiating (57) with respect to η (and evaluating at $\eta = \eta_0$):

$$
\begin{aligned}
-\int\int \begin{pmatrix} 1 \\ x \end{pmatrix} \frac{\partial a(w;\eta_0)}{\partial \eta'} f(y|\, w, x; \eta_0)\, f(x|\, w; \phi_0)\, \mathrm{d}x\mathrm{d}y & \\
-\int\int \begin{pmatrix} x' \\ xx' \end{pmatrix} \frac{\partial b(w;\eta_0)}{\partial \eta'} f(y|\, w, x; \eta_0)\, f(x|\, w; \phi_0)\, \mathrm{d}x\mathrm{d}y & \\
+\left(\int\int \begin{pmatrix} 1 \\ x \end{pmatrix} (y - x'b(w;\eta_0)) \left\{ s_\eta(y|\, w, x; \eta_0) + \left(\frac{\partial \phi(\eta_0)}{\partial \eta'}\right)' \mathbb{S}_\phi(x|\, w) \right\} \right. & \\
\left. \times f(y|\, w, x; \eta_0)\, f(x|\, w; \phi_0)\, \mathrm{d}x\mathrm{d}y \right) &= 0
\end{aligned}
$$

Analogous to the corresponding calculations given in the proof of Theorem 1 we can solve to get

$$
\begin{aligned}
\begin{pmatrix} \frac{\partial a(w;\eta_0)}{\partial \eta'} \\ \frac{\partial b(w;\eta_0)}{\partial \eta'} \end{pmatrix} &= \begin{pmatrix} 1 & -e(w;\phi_0)'\, v(w;\phi_0)^{-1} \\ -v(w;\phi_0)^{-1}\, e(w;\phi_0) & v(w;\phi_0)^{-1} \end{pmatrix} \\
&\times \mathbb{E}\left[\begin{pmatrix} Y - a(W;\eta_0) - X'b(W;\eta_0) \\ X(Y - a(W;\eta_0) - X'b(W;\eta_0)) \end{pmatrix} \right. \\
&\left. \times \left\{ s_\eta(Y|\, W, X; \eta_0) + \left(\frac{\partial \phi(\eta_0)}{\partial \eta'}\right)' \mathbb{S}_\phi(X|\, W) \right\} \right| W = w \Bigg].
\end{aligned}
$$

Plugging the second row of the above expression into (39), which remains unchanged relative to its form in the proof of Theorem 1, we get

$$
\begin{aligned}
\frac{\partial \beta(\eta_0)}{\partial \eta'} &= \mathbb{E}\left[v(W;\phi_0)^{-1} (X - e(W;\phi_0))(Y - a_0(W) - X'b_0(W)) \right. \\
&\left. \times \left\{ s_\eta(Y|\, W, X; \eta_0) + \left(\frac{\partial \phi(\eta_0)}{\partial \eta'}\right)' \mathbb{S}_\phi(X|\, W) \right\} \right] \\
&+ \mathbb{E}\left[b_0(W)\, t_\eta(W) \right]. \qquad (58)
\end{aligned}
$$

Now observe that (17) remains a pathwise derivative. Furthermore (17) continues to lie in the tangent space with its first component playing the role of $s(x, y|\, w) = s(y|\, w, x) + \mathbf{c}\mathbb{S}_\phi(x|\, w)$ and its second component that of $t(w)$. The claim again follows from Theorem 3.1 of Newey (1990).

Detailed calculations for proof of Theorem (2)

Let $m(Z_i, \theta)$ be the $(L+J+1+J+JK+K) \times 1$ vector of moment conditions as defined in the main text. In this appendix we work with the more refined partition of this vector:

$$\begin{aligned} m_1(X_i, W_i, \phi) =& \mathbb{S}_\phi\left(X_i | W_i; \phi\right) && (59)\\ m_2(W_i, \mu_W) =& W_i - \mu_W && (60)\\ m_3(Z_i, \mu_W, \lambda, \beta) =& U_i\left(\mu_W, \lambda, \beta\right) && (61)\\ m_4(Z_i, \mu_W, \lambda, \beta) =& \left(W_i - \mu_W\right) U_i\left(\mu_W, \lambda, \beta\right) && (62)\\ m_5(Z_i, \mu_W, \lambda, \beta) =& \left(\left(W_i - \mu_W\right) \otimes X_i\right) U_i\left(\mu_W, \lambda, \beta\right) && (63)\\ m_5(Z_i, \phi, \mu_W, \lambda, \beta) =& v\left(W;\phi\right)^{-1}\left(X - e\left(W;\phi\right)\right) U_i\left(\mu_W, \lambda, \beta\right) && (64) \end{aligned}$$

where $\theta = (\phi, \mu_W, \lambda', \beta)'$ with $\dim(\theta) = L+J+1+J+JK+K$ as before.

The Jacobian of the moment vector equals

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} & M_{15} & M_{16} \\ M_{21} & M_{22} & M_{23} & M_{24} & M_{25} & M_{26} \\ M_{31} & M_{32} & M_{33} & M_{34} & M_{35} & M_{36} \\ M_{41} & M_{42} & M_{43} & M_{44} & M_{45} & M_{46} \\ M_{51} & M_{52} & M_{53} & M_{54} & M_{55} & M_{56} \\ M_{61} & M_{62} & M_{63} & M_{64} & M_{65} & M_{66} \end{pmatrix}.$$

Considering the first block of columns in M, we have that

$$\underset{L\times L}{M_{11}} = \mathbb{H}\left(\phi_0\right)$$

with $\mathbb{H}(\phi_0)$ equal the $L \times L$ expected Hessian matrix associated with the generalized propensity score log-likelihood (under Assumption 5 we have that $-\mathbb{H}(\phi_0) = \mathbb{E}[\mathbb{S}\mathbb{S}']$). We also have that

$$\underset{J\times L}{M_{21}} = 0,\ \underset{1\times L}{M_{31}} = 0,\ \underset{J\times L}{M_{41}} = 0,\ \underset{JK\times L}{M_{51}} = 0,$$

and, finally

$$\underset{K\times L}{M_{61}} = \mathbb{E}\left[\left(\left[\begin{array}{ccc} -v\left(W,\phi_0\right)^{-1}\frac{\partial v(W,\phi_0)}{\partial\phi_1}v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right) & \cdots \end{array}\right.\right.\right.$$
$$\left.\cdots - v\left(W,\phi_0\right)^{-1}\frac{\partial v\left(W,\phi_0\right)}{\partial\phi_L}v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right)\right]$$
$$\left.\left.+v\left(W,\phi_0\right)^{-1}\frac{\partial e\left(W,\phi_0\right)}{\partial\phi'}\right)U\left(\mu_W,\lambda_*,\beta_0\right)\right].$$

Iterated expectations gives

$$M_{61} = \mathbb{E}\left[\left[\begin{array}{ccc} c_1\left(W,\phi_0\right) & \cdots & c_L\left(W,\phi_0\right) \end{array}\right]\mathbb{C}\left(X,U_*|\,W\right) + d\left(W,\phi_0\right)\mathbb{E}\left[U_*|\,W\right]\right] \tag{65}$$

with $c_l\left(W,\phi\right) = -v\left(W,\phi\right)^{-1}\frac{\partial v(W,\phi)}{\partial\phi_l}v\left(W,\phi\right)^{-1}$ for $l = 1,\ldots,L$ and $d\left(W,\phi\right) = v\left(W,\phi\right)^{-1}\frac{\partial e(W,\phi)}{\partial\phi'}$.

It is useful to develop an alternative expression for (65). Note that

$$\mathbb{E}\left[m_5(Z_i,\phi_0,\mu_W,\lambda_*,\beta_0)\right] = \mathbb{E}\left[v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right)U\left(\mu_W,\lambda_*,\beta_0\right)\right] = 0,$$

is mean zero. A GIME argument, similar to the one used to derive (22) in the main text, therefore gives

$$\mathbb{E}\left[\frac{\partial}{\partial\phi'}\left\{v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right)\right\}U_*\right] = -\mathbb{E}\left[v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right)U_*\mathbb{S}'\right], \tag{66}$$

where we use the fact that $U_* = U\left(\mu_W,\lambda_*,\beta_0\right)$ does not vary with the propensity score parameter, ϕ. We can use (66) to write

$$M_{61} = -\mathbb{E}\left[v\left(W,\phi_0\right)^{-1}\left(X-e\left(W,\phi_0\right)\right)U_*\mathbb{S}'\right].$$

If both Assumptions 5 and 6 hold simultaneously, then $U_* = U_0$ is conditionally mean zero and uncorrelated with X (i.e., $\mathbb{E}\left[U_0|\,W\right] = \mathbb{E}\left[XU_0|\,W\right] = 0$). In this case $M_{61} = 0$ (see Equation (65) above). If Assumption 6 does not hold, then M_{61} may be non-zero.

Turning to the second block of columns in M, we have that

$$\underset{L\times J}{M_{12}} = 0,\ \underset{J\times J}{M_{22}} = -I_J,$$

and also that

$$\begin{aligned}
\underset{1\times J}{M_{32}} =& \mathbb{E}\left[\left\{\gamma_* + (I_J \otimes X)' \delta_*\right\}'\right] \\
\underset{J\times J}{M_{42}} =& \mathbb{E}\left[-I_J U_* + (W - \mu_W)\left\{\gamma_* + (I_J \otimes X)' \delta_*\right\}'\right] \\
=& \mathbb{E}\left[(W - \mu_W)\left\{(I_J \otimes X)' \delta_*\right\}'\right] \\
\underset{JK\times J}{M_{52}} =& \mathbb{E}\left[-(I_J \otimes X) U_* + ((W - \mu_W) \otimes X)\left(\gamma_* + (I_J \otimes X)' \delta_*\right)'\right] \\
\overset{A.6}{=}& \mathbb{E}\left[((W - \mu_W) \otimes X)\left(\gamma_0 + (I_J \otimes X)' \delta_0\right)'\right].
\end{aligned}$$

Note that the second equality after M_{42} does not require Assumption 6 to hold. Even if λ_* does not correctly parameterize the CLP coefficients, it remains true that $U(\mu_W, \lambda_*, \beta_0)$ is mean zero. However $U(\mu_W, \lambda_*, \beta_0)$ may covary with X when Assumption 6 fails. Therefore the second equality after M_{52} *does* require Assumption 6 to hold. The forms of M_{32}, M_{42} and M_{52} determine the effect of sampling uncertainty about the value of μ_W on sampling uncertainty about the value of β_0.

Finally we get

$$\underset{K\times J}{M_{62}} = \mathbb{E}\left[v(W;\phi_0)^{-1}(X - e(W;\phi_0))\left\{\gamma_* + (I_J \otimes X)' \delta_*\right\}'\right]$$

Turning to the third block of columns in M, we have that

$$\underset{L\times 1}{M_{13}} = 0,\ \underset{J\times 1}{M_{23}} = 0,\ \underset{1\times 1}{M_{33}} = -1,$$

and also that

$$\underset{J\times 1}{M_{43}} = -\mathbb{E}\left[(W - \mu_W)\right] = 0,\ \underset{JK\times 1}{M_{53}} = -\mathbb{E}\left[((W - \mu_W) \otimes X)\right]$$

and

$$\underset{K\times 1}{M_{63}} = -\mathbb{E}\left[v(W;\phi_0)^{-1}(X - e(W;\phi_0))\right] = 0.$$

Turning to the fourth block of columns in M, we have that

$$\underset{L\times J}{M_{14}} = 0,\ \underset{J\times J}{M_{24}} = 0,\ \underset{1\times J}{M_{34}} = -\mathbb{E}\left[(W - \mu_W)'\right] = 0,$$

and also that

$$\underset{J\times J}{M_{44}} = -\mathbb{E}\left[(W-\mu_W)(W-\mu_W)'\right] = -\Sigma_{WW}, \ \underset{JK\times J}{M_{54}} = -\mathbb{E}\left[((W-\mu_W)\otimes X)(W-\mu_W)^{'}\right],$$

and finally that

$$\underset{K\times J}{M_{64}} = -\mathbb{E}\left[v(W;\phi_0)^{-1}(X-e(W;\phi_0))(W-\mu_W)^{'}\right] = 0.$$

Turning to the fifth block of columns in M, we have that

$$\underset{L\times JK}{M_{15}} = 0, \ \underset{J\times JK}{M_{25}} = 0,$$

and also that

$$\underset{1\times JK}{M_{35}} = -\mathbb{E}\left[((W-\mu_W)\otimes X)^{'}\right], \ \underset{J\times JK}{M_{45}} = -\mathbb{E}\left[(W-\mu_W)((W-\mu_W)\otimes X)^{'}\right],$$

and also that

$$\underset{JK\times JK}{M_{55}} = -\mathbb{E}\left[((W-\mu_W)\otimes X)((W-\mu_W)\otimes X)^{'}\right],$$

and finally that

$$\underset{K\times JK}{M_{65}} = -\mathbb{E}\left[v(W;\phi_0)^{-1}(X-e(W;\phi_0))((W-\mu_W)\otimes X)^{'}\right] = 0.$$

Turning to the sixth, and final, block of columns in M, we have that

$$\underset{L\times K}{M_{16}} = 0, \ \underset{J\times K}{M_{26}} = 0,$$

and also that

$$\underset{1\times K}{M_{36}} = -\mathbb{E}\left[X^{'}\right], \ \underset{J\times K}{M_{46}} = -\mathbb{E}\left[(W-\mu_W)X^{'}\right], \ \underset{JK\times K}{M_{56}} = -\mathbb{E}\left[((W-\mu_W)\otimes X)X^{'}\right],$$

and finally that

$$\underset{K\times K}{M_{66}} = -\mathbb{E}\left[v(W;\phi_0)^{-1}(X-e(W;\phi_0))X^{'}\right] = -I_K.$$

Marking out the zero and identity terms we have that

$$M = \begin{pmatrix} M_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & -I_J & 0 & 0 & 0 & 0 \\ 0 & M_{32} & -1 & 0 & M_{35} & M_{36} \\ 0 & M_{42} & 0 & M_{44} & M_{45} & M_{46} \\ 0 & M_{52} & M_{53} & M_{54} & M_{55} & M_{56} \\ M_{61} & M_{62} & 0 & 0 & 0 & -I_K \end{pmatrix}.$$

Using the above we have (44), as defined in the appendix to the main paper, equal to

$$\underset{(1+J+JK+K)\times J}{B_1} = \begin{pmatrix} M_{32} \\ M_{42} \\ M_{52} \end{pmatrix} = \mathbb{E}\left[\begin{array}{c} \left\{\gamma_* + (I_J \otimes X)' \delta_*\right\}' \\ (W - \mu_W)\left\{(I_J \otimes X)' \delta_*\right\}' \\ -(I_J \otimes X) U_* + ((W-\mu_W) \otimes X)\left(\gamma_* + (I_J \otimes X)' \delta_*\right)' \end{array} \right].$$

Additional detailed calculations

Equation (46)

To derive the lower-left-hand block of (46) in the Appendix to the main paper we multiply out:

$$-\begin{pmatrix} \mathbb{E}[RR']^{-1} & -\mathbb{E}[RR']^{-1}\mathbb{E}[RX'] \\ 0 & I_K \end{pmatrix}\begin{pmatrix} 0 & -B_1 \\ -M_{61} & -B_2 \end{pmatrix}\begin{pmatrix} -\mathbb{H}(\phi_0)^{-1} & 0 \\ 0 & I_J \end{pmatrix} =$$

$$-\begin{pmatrix} \mathbb{E}[RR']^{-1}\mathbb{E}[RX'] M_{61} & -\mathbb{E}[RR']^{-1}(B_1 - B_2\mathbb{E}[RX']) \\ -M_{61} & -B_2 \end{pmatrix}\begin{pmatrix} -\mathbb{H}(\phi_0)^{-1} & 0 \\ 0 & I_J \end{pmatrix} =$$

$$-\begin{pmatrix} -\mathbb{E}[RR']^{-1}\mathbb{E}[RX'] M_{61}\mathbb{H}(\phi_0)^{-1} & -\mathbb{E}[RR']^{-1}(B_1 - B_2\mathbb{E}[RX']) \\ M_{61}\mathbb{H}(\phi_0)^{-1} & -B_2 \end{pmatrix}.$$

Equation (43)

To derive (43) in the Appendix start by observing that moment (30) in the main text implies the following characterization of λ_0 and β_0 :

$$\begin{bmatrix} \mathbb{E}[RY] \\ \mathbb{E}\left[v_0(W)^{-1}(X - e_0(W))Y\right] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[RR'] & \mathbb{E}[RX'] \\ 0 & I_K \end{bmatrix}\begin{pmatrix} \lambda_0 \\ \beta_0 \end{pmatrix}.$$

After some calculation we get that

$$\begin{aligned}\lambda_0 &= \mathbb{E}\left[RR'\right]^{-1}\mathbb{E}\left[R\left(Y - X'\beta_0\right)\right]\\&= \mathbb{E}\left[RR'\right]^{-1}\mathbb{E}\left[R\left(a_0\left(W\right) + X'\left(b_0\left(W\right) - \beta_0\right)\right) + U_0\right]\\&= \mathbb{E}\left[RR'\right]^{-1}\mathbb{E}\left[R\left(a_0\left(W\right) + X'\left(b_0\left(W\right) - \beta_0\right)\right)\right],\end{aligned}$$

where the last line uses Lemma 1 of the main text.

Equation (50)

To derive equation (50) in the Appendix expand the variance of $b_0\left(W\right) + \Delta_*^{(J)}\left(k^{(J)}\left(W\right) - \mu^{(J)}\right) - \beta_0$ as follows:

$$\begin{aligned}\mathbb{V}\left(b_0\left(W\right) + \Delta_*^{(J)}\left(k^{(J)}\left(W\right) - \mu^{(J)}\right) - \beta_0\right) =& \Delta_*^{(J)}\mathbb{V}\left(k^{(J)}\left(W\right)\right)\left(\Delta_*^{(J)}\right)'\\&+ \mathbb{V}\left(b_0\left(W\right)\right)\\&+ 2\mathbb{E}\left[\left(b_0\left(W\right) - \beta_0\right)\left\{\Delta_*^{(J)}\left(k^{(J)}\left(W\right) - \mu^{(J)}\right)\right\}'\right]\\=& \Delta_*^{(J)}\mathbb{V}\left(k^{(J)}\left(W\right)\right)\left(\Delta_*^{(J)}\right)'\\&+ \mathbb{V}\left(b_0\left(W\right)\right)\\&+ 2\mathbb{E}\left[\left(b_0\left(W\right) - \beta_0\right)\right.\\&\times\left\{b_0\left(W\right) + \Delta_*^{(J)}\left(k^{(J)}\left(W\right) - \mu^{(J)}\right) - \beta_0\right]\\&\times\left. -\left(b_0\left(W\right) - \beta_0\right)\right\}'\Big]\\=& \Delta_*^{(J)}\mathbb{V}\left(k^{(J)}\left(W\right)\right)\left(\Delta_*^{(J)}\right)' - \mathbb{V}\left(b_0\left(W\right)\right)\\&- 2\mathbb{E}\left[\left(b_0\left(W\right) - \beta_0\right)\right.\\&\times\left.\left\{b_0\left(W\right) + \Delta_*^{(J)}\left(k^{(J)}\left(W\right) - \mu^{(J)}\right) - \beta_0\right\}'\right].\end{aligned} \tag{67}$$