IV Weights and Yitzhaki's Theorem

James J. Heckman
University of Chicago

Econ 312, Spring 2021

Digression: Yitzhaki's theorem and extensions

Theorem 1

Assume (Y, X) *i.i.d.* $\quad E(|Y|) < \infty \quad E(|X|) < \infty$

$$\mu_Y = E(Y) \qquad \mu_X = E(X)$$

$E(Y \mid X) = g(X)$

Assume $g'(X)$ *exists and* $E\left(|g'(X)|\right) < \infty$.

Yitzhaki's theorem

Theorem 2 (cont.)

Then,

$$\frac{\mathrm{Cov}(Y,X)}{Var(X)} = \int_{-\infty}^{\infty} g'(t)\,\omega(t)\,dt,$$

where

$$\begin{aligned}\omega(t) &= \frac{1}{Var(X)}\int_{t}^{\infty}(x-\mu_X)\,f_X(x)\,dx \\ &= \frac{1}{Var(X)}E\left(X-\mu_X \mid X>t\right)\Pr\left(X>t\right).\end{aligned}$$

$$Y = \pi X + \eta,$$

$$\pi = \frac{\mathrm{Cov}(Y,X)}{Var(X)}.$$

Proof of Yitzhaki's theorem

Proof.

$$\begin{aligned}\mathrm{Cov}(Y,X) &= \mathrm{Cov}\left(E(Y \mid X), X\right) = \mathrm{Cov}\left(g(X), X\right) \\ &= \int_{-\infty}^{\infty} g(t)(t-\mu_X)\, f_X(t)\, dt\end{aligned}$$

where t is an argument of integration.

cont.

Integration by parts:

$$
\begin{aligned}
\mathrm{Cov}(Y, X) &= g(t) \int_{-\infty}^{t} (x - \mu_X)\, f_X(x)\, dx \Big|_{-\infty}^{\infty} \\
&\qquad - \int_{-\infty}^{\infty} g'(t) \int_{-\infty}^{t} (x - \mu_X)\, f_X(x)\, dx\, dt \\
&= \int_{-\infty}^{\infty} g'(t) \int_{t}^{\infty} (x - \mu_X)\, f_X(x)\, dx\, dt,
\end{aligned}
$$

since $E(X - \mu_X) = 0$.

THE UNIVERSITY OF CHICAGO

cont.

Therefore,

$$\mathrm{Cov}(Y,X) = \int_{-\infty}^{\infty} g'(t)\, E\left(X - \mu_X \mid X > t\right) \Pr\left(X > t\right)\, dt.$$

$\therefore$ Result follows with

$$\omega(t) = \frac{1}{Var(X)} E\left(X - \mu_X \mid X > t\right) \Pr\left(X > t\right)$$

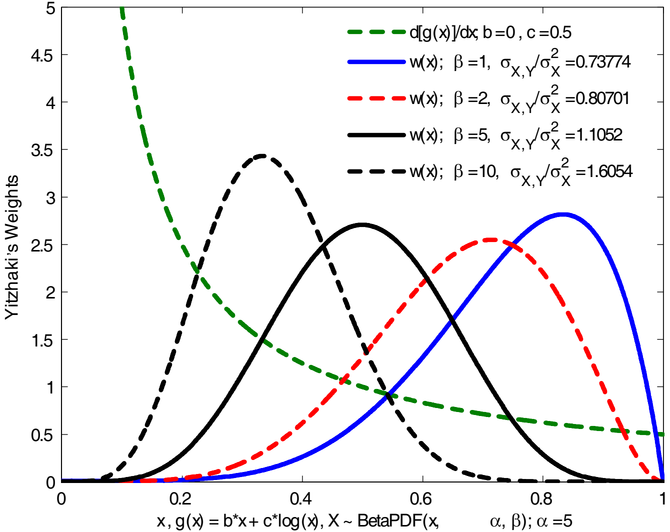
□

THE UNIVERSITY OF CHICAGO

- Weights positive.
- Integrate to one (use integration by parts formula).
- $= 0$ when $t \to \infty$ and $t \to -\infty$.
- Weight reaches its peak at $t = \mu_X$, if f_X has density at $x = \mu_X$:

$$\begin{aligned} \frac{d}{dt} \int_t^{\infty} (x - \mu_X)\, f_X(x)\, dx\, dt &= -(t - \mu_X) f_X(t) \\ &= 0 \quad \text{at } t = \mu_X. \end{aligned}$$

Yitzhaki's weights for $X \sim$ BetaPDF(x, α, β)



$$E(Y|X=x)=g(x)\Rightarrow\frac{Cov(X,Y)}{Var(X)}=\int_{-\infty}^{\infty}g'(t)w(t)dx$$

$$w(t)=\frac{1}{Var(X)}E(X|X>t)\cdot\Pr(X>t)$$

$$\mathbf{X}\sim BetaPDF(x,\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)};\ \alpha=5;$$

$$\mathbf{g}(\mathbf{x})=\mathbf{0.5}\cdot\mathbf{x}+\mathbf{0.5}\cdot\log(\mathbf{X})$$

THE UNIVERSITY OF CHICAGO

- Can apply Yitzhaki's analysis to the treatment effect model

$$Y = \alpha + \beta D + \varepsilon$$

- $P(Z)$, the propensity score is the instrument:

$$E\left(Y \mid Z = z\right) = E\left(Y \mid P(Z) = p\right)$$

$$
\begin{aligned}
E\left(Y \mid P(Z)=p\right) &= \alpha + E\left(\beta D \mid P(Z)=p\right) \\
&= \alpha + E\left(\beta \mid D=1, P(Z)=p\right) p \\
&= \alpha + E\left(\beta \mid P(Z) > U_D, P(Z)=p\right) p \\
&= \alpha + E\left(\beta \mid p > U_D\right) p \\
&= \alpha + \underbrace{\int \beta \int_0^p f\left(\beta, u_D\right)\, du_D}_{g(p)}
\end{aligned}
$$

- Derivative with respect to p is MTE.
- $g'(p) =$MTE and weights as before.

- Under uniformity,

$$\frac{\partial E\left(Y \mid P\left(Z\right)=p\right)}{\partial p} = E\left(Y_1 - Y_0 \mid U_D = u_D\right) = \Delta^{MTE}\left(u_D\right).$$

- More generally, it is LIV $= \frac{\partial E(Y|P(Z)=p)}{\partial p}$.
- Yitzhaki's result does not rely on uniformity; true of any regression of Y on P.
- Estimates a weighted net effect.
- The expression can be generalized.
- It produces Heckman-Vytlacil weights.

The Heckman-Vytlacil weight as a Yitzhaki weight

Consider a general function of Z, $J(Z)$.

Proof.

$$
\begin{aligned}
\operatorname{Cov}(J(Z), Y) &= E\left(Y \cdot \widetilde{J}\right) = E\left(E(Y \mid Z) \cdot \widetilde{J}(Z)\right) \\
&= E\left(E(Y \mid P(Z)) \cdot \widetilde{J}(Z)\right) \\
&= E\left(g(P(Z)) \cdot \widetilde{J}(Z)\right).
\end{aligned}
$$

$$
\begin{gathered}
\widetilde{J} = J(Z) - E(J(Z) \mid P(Z) \geq u_D), \\
E(Y \mid P(Z)) = g(P(Z)).
\end{gathered}
$$

cont.

$$\begin{aligned}\mathrm{Cov}\left(J\left(Z\right), Y\right) &= \int_0^1 \int_{\underline{J}}^{\overline{J}} g\left(u_D\right) \widetilde{j} f_{P,J}\left(u_D, j\right) \mathrm{d}j \mathrm{d}u_D \\ &= \int_0^1 g\left(u_D\right) \int_{\underline{J}}^{\overline{J}} \widetilde{j} f_{P,J}\left(u_D, j\right) \mathrm{d}j \mathrm{d}u_D.\end{aligned}$$

cont.

Use integration by parts:

$$
\begin{aligned}
&\mathrm{Cov}\left(J\left(Z\right), Y\right) \\
&= g\left(u_D\right) \int_0^{u_D} \int_{\underline{J}}^{\overline{J}} \widetilde{j} f_{P,J}\left(p, j\right)\, \mathrm{d}j \mathrm{d}p \Big|_0^1 \\
&\qquad - \int_0^1 g'\left(u_D\right) \int_0^{u_D} \int_{\underline{J}}^{\overline{J}} \widetilde{j} f_{P,J}\left(p, j\right)\, \mathrm{d}j \mathrm{d}p \mathrm{d}u_D \\
&= \int_0^1 g'\left(u_D\right) \int_{u_D}^{1} \int_{\underline{J}}^{\overline{J}} \widetilde{j} f_{P,J}\left(p, j\right)\, \mathrm{d}j \mathrm{d}p \mathrm{d}u_D \\
&= \int_0^1 g'\left(u_D\right) E\left(\widetilde{J}(Z) \mid P\left(Z\right) \geq u_D\right) \Pr\left(P\left(Z\right) \geq u_D\right) \mathrm{d}u_D.
\end{aligned}
$$

cont.

Thus:

$$g'(u_D) = \left.\frac{\partial E(Y \mid P(Z) = p)}{\partial P(Z)}\right|_{p=u_D} = \Delta^{\text{MTE}}(u_D).$$

□

- Under our assumptions the Yitzhaki weights and ours are equivalent.

$$\begin{aligned} & \operatorname{Cov}\left(J\left(Z\right), Y\right) \\ &= \int_0^1 \Delta^{\text{MTE}}(u_D) E(J(Z) - E(J(Z)) | P(Z) \geq u_D) \Pr(P(Z) \geq u_D) du_D. \end{aligned} \tag{1}$$

- Using (1),

$$\begin{aligned} \operatorname{Cov}\left(J\left(Z\right), Y\right) &= E\left(Y \cdot \tilde{J}\right) = E\left(E\left(Y \mid Z\right) \cdot \tilde{J}(Z)\right) \\ &= E\left(E\left(Y \mid P\left(Z\right)\right) \cdot \tilde{J}(Z)\right) \\ &= E\left(g\left(P\left(Z\right)\right) \cdot \tilde{J}(Z)\right). \end{aligned}$$

THE UNIVERSITY OF CHICAGO

- The third equality follows from index sufficiency and $\tilde{J} = J(Z) - E(J(Z) \mid P(Z) \geq u_D)$, where $E(Y \mid P(Z)) = g(P(Z))$.
- Writing out the expectation and assuming that $J(Z)$ and $P(Z)$ are continuous random variables with joint density $f_{P,J}$ and that $J(Z)$ has support $\left[\underline{J}, \overline{J}\right]$,

$$\begin{aligned}
\operatorname{Cov}(J(Z), Y) &= \int_0^1 \int_{\underline{J}}^{\overline{J}} g(u_D)\, \tilde{j} f_{P,J}(u_D, j)\, \mathrm{d}j \mathrm{d}u_D \\
&= \int_0^1 g(u_D) \int_{\underline{J}}^{\overline{J}} \tilde{j} f_{P,J}(u_D, j)\, \mathrm{d}j \mathrm{d}u_D .
\end{aligned}$$

- Using an integration by parts argument as in Yitzhaki (1989) and as summarized in Heckman, Urzua, Vytlacil (2006), we obtain

$$
\begin{aligned}
&\operatorname{Cov}(J(Z), Y) \\
&= g(u_D) \int_0^{u_D} \int_{\underline{J}}^{\overline{J}} \tilde{j} f_{P,J}(p, j) \, \mathrm{d}j \mathrm{d}p \Big|_0^1 \\
&\qquad - \int_0^1 g'(u_D) \int_0^{u_D} \int_{\underline{J}}^{\overline{J}} \tilde{j} f_{P,J}(p, j) \, \mathrm{d}j \mathrm{d}p \mathrm{d}u_D \\
&= \int_0^1 g'(u_D) \int_{u_D}^1 \int_{\underline{J}}^{\overline{J}} \tilde{j} f_{P,J}(p, j) \, \mathrm{d}j \mathrm{d}p \mathrm{d}u_D \\
&= \int_0^1 g'(u_D) E\left(\tilde{J}(Z) \mid P(Z) \geq u_D\right) \Pr(P(Z) \geq u_D) \, \mathrm{d}u_D,
\end{aligned}
$$

which is then exactly the expression given in (1), where

$$
g'(u_D) = \left. \frac{\partial E(Y \mid P(Z) = p)}{\partial P(Z)} \right|_{p=u_D} = \Delta^{\mathrm{MTE}}(u_D).
$$

Separable choice model

$$\Delta_J^{IV} = \int_0^1 \Delta^{MTE}(u_D)\,\omega_{IV}^J(u_D)\,du_D \tag{2}$$

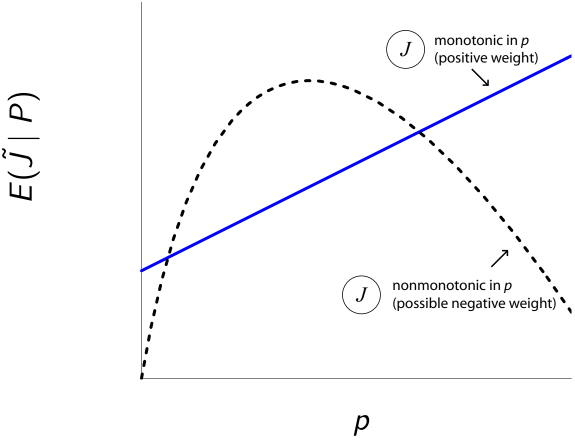
$$\omega_{IV}^J(u_D) = \frac{E\left(J(Z) - \bar{J}(Z) \mid P(Z) > u_D\right)\Pr\left(P(Z) > u_D\right)}{\mathrm{Cov}\left(J(Z), D\right)}. \tag{3}$$

$J(Z)$ and $P(Z)$ do not have to be continuous random variables.

Functional forms of $P(Z)$ and $J(Z)$ are general.

- Dependence between $J(Z)$ and $P(Z)$ gives shape and sign to the weights.
- If $J(Z) = P(Z)$, then weights obviously non-negative.
- If $E(J(Z) - \bar{J}(Z) \mid P(Z) \geq u_D)$ not monotonic in u_D, weights can be negative.

$$\tilde{J} = J - E(J)$$



Therefore, with positive (or negative) regression, can get negative IV weight.

When $J(Z) = P(Z)$, weight (3) follows from Yitzhaki (1989).

- He considers a regression function $E(Y \mid P(Z) = p)$.
- Linear regression of Y on P identifies

$$\beta_{Y,P} = \int_0^1 \left[\frac{\partial E(Y \mid P(Z) = p)}{\partial p} \right] \omega(p)\, dp,$$

$$\omega(p) = \frac{\int_p^1 (t - E(P))\, dF_P(t)}{Var(P)}.$$

- This is the weight (3) when P is the instrument.
- This expression **does not** require uniformity or monotonicity for the model; consistent with 2-way flows.

Understanding the structure of the IV weights

Recapitulate:

$$\Delta_{\text{IV}}^{J} = \int \Delta^{\text{MTE}}(u_D)\,\omega_{\text{IV}}^{J}(u_D)\,du_D$$

$$\omega_{\text{IV}}^{J}(u_D) = \frac{\int (j - E(J(Z))) \int_{u_D}^{1} f_{J,P}(j,t)\ dt\, dj}{Cov(J(Z), D)} \tag{4}$$

- The weights are always positive if $J(Z)$ is monotonic in the scalar Z.
- In this case $J(Z)$ and $P(Z)$ have the same distribution and $f_{J,P}(j,t)$ collapses to a single distribution.

- The possibility of negative weights arises when $J(Z)$ is not a monotonic function of $P(Z)$.
- It can also arise when there are two or more instruments, and the analyst computes estimates with only one instrument or a combination of the Z instruments that is not a monotonic fuction of $P(Z)$ so that $J(Z)$ and $P(Z)$ are not perfectly dependent.

- The weights can be constructed from data on (J, P, D).
- Data on $(J(Z), P(Z))$ pairs and $(J(Z), D)$ pairs (for each X value) are all that is required.