Duality in Discrete-Choice Models

Thomas Wiemann
University of Chicago

TA Discussion # 6
Econ 31740

February 21, 2022

Overview

Key steps of Chiong et al. (2016) are:

1. Identification of choice-specific payoffs using convex duality.
   - ▷ Convex analysis as a useful tool for (partial) identification.
2. Show that choice-specific payoffs correspond to Lagrange multipliers of an Optimal Transport problem.
3. Propose a linear programming estimator for discrete choice models based on discretized version of the optimal transport problem.
   - ▷ Computational advantages over simulated ML.

Contributions are in steps 1. and 3., step 2. is a contribution of Galichon and Salanié (2020).

In contrast to Chiong et al. (2016), I will focus on *static* discrete choice.

Discrete Choice Model

Consider the classical discrete choice problem where heterogeneous agents choose an alternative $y \in \mathcal{Y} := \{1, \ldots, J\}$ according to

$$y = \underset{y \in \mathcal{Y}}{\arg\max}\ w_y + \varepsilon_y, \tag{1}$$

where w_y is the systematic utility of choice y shared across all agents, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J) \sim Q$ is the vector of latent utility shocks.

Define

$$Y(w, \varepsilon) := \underset{y \in \mathcal{Y}}{\arg\max}\ w_y + \varepsilon_y, \tag{2}$$

where $w := (w_j)_{j=1}^J$ and $\varepsilon := (\varepsilon_j)_{j=1}^J$, and let

$$p_y := E_Q\left[Y(w, \varepsilon) = y\right] = P_Q(w_y + \varepsilon_y \geq w_j + \varepsilon_j, \forall j \neq y). \tag{3}$$

Discrete Choice Model (Contd.)

Suppose:

- the choice probabilities $(p_y)_{y \in \mathcal{Y}}$ are observed;
- the distribution Q is known to the researcher (e.g., Q is T1EV).

The identified set of the choice-specific utilities $(w_y)_{y \in \mathcal{Y}}$ is

$$\mathcal{I}(p) := \left\{ w \in \mathbb{R}^J \,|\, p_y = E_Q \left[Y(w, \varepsilon) = y \right], \forall y \in \mathcal{Y} \right\}. \tag{4}$$

Informally: *Given the choice probabilities and the distribution of the latent utility shocks, what are the systematic utilities that are compatible with the discrete choice problem* (1)*?*

Discrete Choice Model (Contd.)

Define the ex-ante expected utility (or social surplus) as

$$G(w) := E_Q\left[\max_{y\in\mathcal{Y}}\ w_y + \varepsilon_y\right]. \tag{5}$$

If $E[\varepsilon] < \infty$, then $\forall w < \infty$, we have $|G(w)| < \infty$ and

$$\begin{aligned}
\frac{\partial G(w)}{\partial w_y} &= \frac{\partial}{\partial w_y}\int \max_{y\in\mathcal{Y}}\{w_y + \varepsilon_y\}dQ(\varepsilon) \\
&= \int \frac{\partial}{\partial w_y}\max_{y\in\mathcal{Y}}\{w_y + \varepsilon_y\}dQ(\varepsilon) \\
&= \int \mathbb{1}\{w_y + \varepsilon_y \geq w_j + \varepsilon_j, \forall j \neq y\}dQ(\varepsilon) \\
&= p_y
\end{aligned} \tag{6}$$

Equation (6) is the Williams-Daly-Zachary Theorem (see, e.g., Proposition 1 in Chiong et al. 2016 or Theorem 3.1 in Rust 1994).

Convex Duality for Discrete Choice Models

Equation (6) provides a mapping from the systematic utilities w to the choice probabilities p.

This is the "reverse" problem to the identification question posed in (4), where w is unobserved to the researcher and p is observed.

Convex duality relates Equation (6) with (4).

The next two slides review basic definitions and results from convex analysis. Exposition is taken from Çınlar and Vanderbei (2013).

Convex Duality for Discrete Choice Models (Contd.)

Definition 1 (Convex conjugate)

Let $f : \mathbb{R}^n \to \mathbb{R}^* := \mathbb{R} \cup \{+\infty\}$ be convex. Its *convex conjugate* (or Legendre transform) is the function $f^* : \mathbb{R}^n \to \mathbb{R}^*$ defined by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} \xi^\top x - f(x), \quad \forall \xi \in \mathbb{R}^n.$$

Definition 2 (Subdifferential)

Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be convex. The *subdifferential* of f at x is

$$\begin{aligned} \partial f(x) &:= \left\{ \xi \in \mathbb{R}^n | f(x') \geq f(x) + \xi^\top (x' - x), \, \forall x' \in \mathbb{R}^n \right\} \\ &= \left\{ \xi \in \mathbb{R}^n | \xi^\top x - f(x) \geq \xi^\top x' - f(x'), \, \forall x' \in \mathbb{R}^n \right\}. \end{aligned}$$

Convex Duality for Discrete Choice Models (Contd.)

Theorem 1 (Legrende-Fenchel equality)

Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be convex and fix $x \in \mathbb{R}^n$. Then the following are equivalent:

(a) $\xi \in \partial f(x)$.

(b) $x = \arg\max_{x' \in \mathbb{R}^n} \xi^\top x' - f(x')$.

(c) $f(x) + f^*(\xi) = \xi^\top x$.

(d) $x \in \partial f^*(\xi)$.

(e) $\xi = \arg\max_{\xi' \in \mathbb{R}^n} \xi'^\top x - f^*(\xi')$.

If $E[\varepsilon] < \infty$, then $G(w) = E_Q\left[\max_{y \in \mathcal{Y}} w_y + \varepsilon_y\right]$ is convex so that we may apply Theorem 1.

Convex Duality for Discrete Choice Models (Contd.)

The convex conjugate of G is given by

$$\begin{aligned} G^*(p) &= \sup_{w \in \mathbb{R}^J} p^\top w - G(w) \\ &= \sup_{w \in \mathbb{R}^J} p^\top w - E_Q\left[\max_{y \in \mathcal{Y}} w_y + \varepsilon_y\right]. \end{aligned} \tag{7}$$

Notice that $G^*(p) = \infty$ if $p \notin \{p \in \mathbb{R}^J |\, p_y \geq 0, \forall y \in \mathcal{Y}, \sum_{y \in \mathcal{Y}} p_y \leq 1\}$.

By application of Theorem 1, we then immediately obtain Theorem 2:

Theorem 2 (Theorem 1 in Chiong et al., 2016)

Assume $E_Q[\varepsilon] < \infty$. Then

$$p \in \partial G(w) \quad \Leftrightarrow \quad w \in \partial G^*(p).$$

Identification of the Logit Model using Convex Analysis

Suppose Q is $\mathrm{T1EV}$, then

$$G(w) = \log\left(\sum_{y\in\mathcal{Y}} \exp\left(w_y\right)\right) + \gamma, \tag{8}$$

and

$$G^*(p) = \sup_{w\in\mathbb{R}^J} p^\top w - G(w) = \sum_{y\in\mathcal{Y}} p_y \log p_y - \gamma, \tag{9}$$

whenever $p \in \Delta^J := \{p \in \mathbb{R}^J |\, p_y \geq 0, \forall y \in \mathcal{Y}, \sum_{y\in\mathcal{Y}} p_y \leq 1\}$ and $G^*(p) = \infty$ otherwise, where $\gamma \approx 0.5772$ is Euler's constant.

The subdifferential of G^* at $p \in \Delta^J$ is

$$\partial G^*(p) = \left\{ w \in \mathbb{R}^J |\, \sum_{y\in\mathcal{Y}} p'_y \left(\log p'_y - w_y\right) \geq \sum_{y\in\mathcal{Y}} p_y \left(\log p_y - w_y\right), \forall p' \in \Delta^J \right\}.$$

Hence

$$w \in \partial G^*(p) \quad \Leftrightarrow \quad \exists k \in \mathbb{R} : w_y = \log p_y - k, \forall y \in \mathcal{Y}.$$

An Indeterminacy Problem

Let $p \in \Delta^J$ and $w \in \partial G^*(p)$. Then $\forall k \in \mathbb{R}$, $(w + k\mathbf{1}_J) \in \partial G^*(p)$.

To see this, conjecture

$$\begin{aligned} G^*(p') &\geq G^*(p) + \left[(w+k)^\top p' - (w + k\mathbf{1}_J)^\top p\right] \\ &= G^*(p) + w^\top (p' - p) + k(\mathbf{1}_J^\top p' - \mathbf{1}_J^\top p), \end{aligned} \tag{10}$$

where either $p' \in \Delta^J \Rightarrow k(\mathbf{1}_J^\top p' - \mathbf{1}_J^\top p) \leq 0$ or $p' \notin \Delta^J \Rightarrow G^*(p') = \infty$, so that the conjecture holds $\forall p' \in \mathbb{R}^J, k \in \mathbb{R}$ whenever $w \in \partial G^*(p)$.

Differences in w_y are important. Fix their levels via

$$w^0 : \; G(w^0) = 0. \tag{11}$$

(We will see that this is indeed a normalization.)

Identification via Convex Duality

Theorem 3 (Theorem 2 in Chiong et al., 2016)

Assume the distribution Q of the latent utility shocks ε is such that $E_Q[\varepsilon] < \infty$ and the distribution of the vector $(\varepsilon_y - \varepsilon_1)_{y \neq 1}$ has full support. Let $p \in \text{Int}\left(\{p \in \mathbb{R}_+^J | \sum_j p_j = 1\}\right)$. Then, for a given Q, there exists a unique $w^0 \in \partial G^*(p)$ such that $G(w^0) = 0$.

sketched proof

Full support assumption common in the literature (e.g., Rust, 1994).

- all $y \in \mathcal{Y}$ have positive probability in all choice sets

Identification via Convex Duality (Contd.)

Theorem 4 (Theorem 3 in Chiong et al., 2016)

Let $k \in \mathbb{R}$ and maintain the assumptions of Theorem 3 on Q and p. The set of conditions

$$w \in \partial G^*(p) \quad \text{and} \quad G(w) = k, \tag{12}$$

is equivalent to

$$w_y = w_y^0 + k, \forall y \in \mathcal{Y}. \tag{13}$$

sketched proof

w^0 is a convenient reference point $\forall w \in \partial G^*(p)$.

Computation using Optimal Transport

For some Q, $\partial G^*(p)$ can be easily characterized (e.g., when Q is T1EV).

If simple characterizations don't exists, part (b) of Theorem 1 provides a constructive approach for it's computation.

By Theorem 1 (b), we have $w \in \partial G^*(p)$ if and only if

$$w = \underset{w' \in \mathbb{R}^J}{\arg\max}\, p^\top w' - E_Q \left[\max_{y \in \mathcal{Y}} w_y + \varepsilon_y \right]. \tag{14}$$

The next theorem states that (14) is equivalent to the dual of an optimal transport problem.

Computation using Optimal Transport (Contd.)

Theorem 5 (Proposition 2 in Chiong et al., 2016)

Maintain the assumptions of Theorem 3 on Q and p. Then

$$G^*(p) = \sup_{w,z\colon w_y + z(\varepsilon) \leq c(y,\varepsilon)} E_p\left[w_Y\right] + E_Q\left[z(\varepsilon)\right], \tag{15}$$

where $c(y,\varepsilon) = -\varepsilon_y$, $w \in \mathbb{R}^J$, and $z(\cdot)$ is a Q-measurable random variable. By Monge-Kantorovich duality, (15) *coincides with its dual*

$$G^*(p) = \min_{\pi\colon Y \sim p, \varepsilon \sim Q} E_\pi\left[c(Y,\varepsilon)\right]. \tag{16}$$

Further, $w \in \partial G^*(p)$ if and only if there exists z such that (w, z) solves (15)*. Finally, $w^0 \in \partial G^*(p)$ and $G(w^0) = 0$ if and only if there exists z such that (w^0, z) solves* (15) *and z is such that $E_Q[z(\varepsilon)] = 0$*

reformulation of (14) to (15)

Another Indeterminacy Problem

Theorem 5 requires that Q satisfies a full support assumption.

- Q cannot be discrete
- (15) and (16) are *infinitely* dimensional linear programming problems
- computationally challenging without closed form expression

The next theorem defines the identified set of systematic utilities given choice probabilities in settings when Q does not satisfy the full support.

- useful for estimands of discretized versions of (15) and (16)

Another Indeterminacy Problem (Contd.)

Theorem 6 (Theorem 4 in Chiong et al., 2016)

Assume the distribution Q of the latent utility shocks ε is such that $E_Q[\varepsilon] < \infty$, and let $p \in \mathrm{Int}\left(\{p \in \mathbb{R}_+^J | \sum_j p_j = 1\}\right)$. $\mathcal{I}(p)$ is the set of w such that there exists a z such that (w, z) is a solution to (15). Therefore,

$$\mathcal{I}(p) = \left\{ w \in \mathbb{R}^J | \, \exists z, \, w_y + z_\varepsilon \leq c(y, \varepsilon), \, E_p[w_Y] + E_Q[z_\varepsilon] = G^*(p) \right\},$$

and

$$\mathcal{I}_0(p) = \left\{ w \in \mathbb{R}^J | \, \exists z, \, w_y + z_\varepsilon \leq c(y, \varepsilon), \, E_p[w_Y] = G^*(p), \, E_Q[z_\varepsilon] = 0 \right\}.$$

Estimation

Let $\hat{Q}$ be a discrete approximation to Q.

- $\hat{Q}$ discrete uniform over S iid samples from Q: $\{\varepsilon_y^s\}_{y\in\mathcal{Y},s\in\{1,\dots,S\}}$.

Then, the discretized analogue to (15) is

$$\begin{aligned} \max_{w\in\mathbb{R}^J, z\in\mathbb{R}^S} \quad & \sum_{y\in\mathcal{Y}} p_y w_y + \frac{1}{S}\sum_{s=1}^{S} z_s \\ \text{s.t.} \quad & w_y + z_s \le -\varepsilon_y^s, \qquad \forall y\in\mathcal{Y}, s\in\{1,\dots,S\}, \end{aligned} \tag{17}$$

where $\{p_y\}_{y\in\mathcal{Y}}$ is known (or estimated).

Let w_n^0 denote the solution to (17) with estimated choice probabilities.

- Theorem 5 of Chiong et al. (2016) gives conditions for $w_n^0 \xrightarrow{a.s.} w^0$.

Alternatively, w can be estimated as the Lagrange multipliers to the discretized primal problem.

Estimation (Contd.)

Theorem 6 suggests a straightforward way of computing upper and lower bounds for each w_y^0 :

1. Construct the discrete approximation $\hat{Q}$ to Q by randomly sampling S iid draws from Q.
2. Solve the linear program in (17) to obtain the objective value $G^*(p)$.
3. Compute a lower bound for w_y^0 via the linear program

$$\begin{aligned} \min_{w\in\mathbb{R}^J, z\in\mathbb{R}^S} \quad & w_y \\ \text{s.t.} \quad & w_y + z_s \leq -\varepsilon_y^s, \qquad \forall y \in \mathcal{Y}, s \in \{1, \ldots, S\}, \\ & \sum_{y\in\mathcal{Y}} p_y w_y = G^*(p), \\ & \frac{1}{S}\sum_{s=1}^{S} z_s = 0. \end{aligned} \tag{18}$$

An upper bound may be calculated analogously using max instead.

Empirical Illustrations

Chiong et al. (2016) conduct a simulation exercise and consider an empirical application.

Simulation exercise using dynamic resource extraction model:

- ▷ illustrates finite sample performance (when $\hat{p}_n$ is used)
- ▷ suggests indeterminacy problem practically not important
- ▷ highlights computational advantages over simulated ML

Empirical application to the data of Rust (1987):

- ▷ dynamic model of bus engine replacement
- ▷ state space $\mathcal{X}$ is 30 states of discretized bus mileage
- ▷ estimate w independently for each state $x \in \mathcal{X}$ where $\hat{p}_n^x > \mathbf{0}$

I implement (17) and (18) in Julia (link) to run a simple MC exercise.

Discussion and Extensions

Summary:

- Convex duality appears to be a promising tool for analyzing (partial) identification of discrete choice models.
- Computationally attractive approach, allows to consider different distributions Q, not just those with convenient closed forms.

Extensions:

- Often interested in characterizing w_y as a function of observable market and product characteristics. The approach of Chiong et al. (2016) does not immediately allow for smoothing across characteristics and instead requires estimation *per market*.
- Inference approach discussed in Hsieh et al. (2022).

References

Chiong, K. X., Galichon, A., and Shum, M. (2016). Duality in dynamic discrete-choice models. *Quantitative Economics*, 7(1):83–115.

Çınlar, E. and Vanderbei, R. J. (2013). *Real and Convex Analysis*. Springer Science & Business Media.

Galichon, A. and Salanié, B. (2020). Cupid's invisible hand: Social surplus and identification in matching models. *SSRN working paper No 1804623*.

Hsieh, Y.-W., Shi, X., and Shum, M. (2022). Inference on estimators defined by mathematical programming. *Journal of Econometrics*, 226(2):248–268.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, pages 999–1033.

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143.

Sketched Proof of Theorem 3

back

1. Choose $\tilde{w} \in \partial G^*(p)$ and let $w_y = \tilde{w}_y - G(\tilde{w})$. Note that $G(w) = E\left[\max_y \tilde{w}_y - G(\tilde{w}) + \varepsilon_y\right] = G(\tilde{w}) - G(\tilde{w}) = 0$, and $w \in \partial G^*(p)$. Then $p = \partial G(w)$ by Theorem 2.

2. To show uniqueness, suppose that $\exists w \neq w' : \ G(w) = G(w') = 0$ and $p \in \partial G(w)$ and $p \in \partial G(w')$. Then $\exists y_0 \neq y_1 : \ w_{y_0} - w_{y_1} \neq w'_{y_0} - w'_{y_1}$.

3. WLOG, consider $w_{y_0} - w_{y_1} > w'_{y_0} - w'_{y_1}$. Define

$$\Gamma := \left\{ \varepsilon \in \operatorname{supp} Q \,\middle|\, \begin{array}{l} w_{y_0} - w_{y_1} > \varepsilon_{y_1} - \varepsilon_{y_0} > w'_{y_0} - w'_{y_1} \\ w_{y_0} + \varepsilon_{y_0} > \max\limits_{y \neq y_0, y_1} w_y + \varepsilon_y \\ w'_{y_1} + \varepsilon_{y_1} > \max\limits_{y \neq y_0, y_1} w'_y + \varepsilon_y \end{array} \right\}. \tag{19}$$

Sketched Proof of Theorem 3 (Contd.)

back

4. Note that $\forall \varepsilon \in \Gamma$, it holds that $Y(w, \varepsilon) = y_0$ and $Y(w', \varepsilon) = y_1$. Because of the full support assumption, $P_Q(\varepsilon \in S) > 0$.

5. Let $\bar{w} = \frac{w+w'}{2}$. Because $p \in \partial W^{(}w)$ and $p \in \partial W^{(}w')$, W is linear on $[w, w']$; combining with $G(w) = G(w') = 0$ we have $G(\bar{w}) = 0$.

6. Thus

$$\begin{aligned} 0 &= E\left[\bar{w}_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] \\ &= \frac{1}{2}E\left[w_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] + \frac{1}{2}E\left[w'_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] \\ &\leq \frac{1}{2}E\left[w_{Y(w,\varepsilon)} + \varepsilon_{Y(w,\varepsilon)}\right] + \frac{1}{2}E\left[w'_{Y(w',\varepsilon)} + \varepsilon_{Y(w',\varepsilon)}\right] \\ &= \frac{1}{2}\left(G(w) + G(w')\right) = 0, \end{aligned} \tag{20}$$

where we used $w_{Y(w,\varepsilon)} + \varepsilon_{Y(w,\varepsilon)} \geq w_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}$.

Sketched Proof of Theorem 3 (Contd.)

back

7. It follows from $w^l_{Y(w^l,\varepsilon)} + \varepsilon_{Y(w^l,\varepsilon)} \geq w^l_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}$, $\forall w^l \in \{w, w'\}$ and (20) that the equality holds term by term:

$$\begin{aligned} w_{Y(w,\varepsilon)} + \varepsilon_{Y(w,\varepsilon)} &\geq w_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}, \\ w'_{Y(w',\varepsilon)} + \varepsilon_{Y(w',\varepsilon)} &\geq w'_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}. \end{aligned} \tag{21}$$

8. Take $\varepsilon \in \Gamma$. Then $y_0 = Y(w,\varepsilon) = Y(\bar{w},\varepsilon) = Y(w',\varepsilon) = y_1$, which is the desired contradiction.

9. Hence $w = w'$ and uniqueness follows.

Sketched Proof of Theorem 4

back

Recall $G(w^0) = 0$ by definition and note that

$$\partial G\left(w - G(w)\right) = \partial G(w). \tag{22}$$

Then, by uniqueness of w^0 in Theorem 3, it follows that

$$w^0 = w - G(w). \tag{23}$$

Define $z(\varepsilon) := -\min_{y\in\mathcal{Y}}\{-w_y - \varepsilon_y\}$ and introduce the constraint

$$\begin{aligned} & \max_{y\in\mathcal{Y}}\{w_y + \varepsilon_y\} \geq w_y + \varepsilon_y, && \forall y \in \mathcal{Y} \\ \Leftrightarrow \quad & -z(\varepsilon) \geq w_y + \varepsilon, && \forall y \in \mathcal{Y} \\ \Leftrightarrow \quad & c(y,\varepsilon) \geq w_y + z(\varepsilon), && \forall y \in \mathcal{Y}, \end{aligned} \tag{24}$$

where $c(y,\varepsilon) := -\varepsilon_y$.

Then w satisfies Equation (14) if and only if there exists a Q-measurable random variable z such that

$$\begin{aligned} (w,z) = \arg\max_{w',z}\, & p^\top w' - E_Q\left[z(\varepsilon)\right] \\ \text{s.t.} \quad & c(y,\varepsilon) \geq w_y + z(\varepsilon), \quad \forall y \in \mathcal{Y}. \end{aligned} \tag{25}$$

Discretized Primal Problem

back

Note that $\{w_y\}$ are the Lagrange multipliers to the first set of constraints of the discretized primal problem given by

$$\begin{aligned} \min_{\pi \in \mathbb{R}_+^{J \times S}} \quad & -\sum_{y,s} \pi_{ys} \varepsilon_y^s \\ \text{s.t.} \quad & \sum_{s=1}^{S} \pi_{ys} = p_y, && \forall y \in \mathcal{Y}, \\ & \sum_{y \in \mathcal{Y}} \pi_{ys} = \frac{1}{S}, && \forall s \in \{1, \ldots, S\}. \end{aligned} \tag{26}$$

The optimal $\{w_y\}_{y \in \mathcal{Y}}$ can thus be obtained using either (17) or (26).

Chiong et al. (2016) note that they implement the primal problem (26) for computation, not its dual (17). Using modern solvers, it likely doesn't make a difference (?).

Simulation Exercise

back

Consider a simple simulation setup where Q is T1EV and

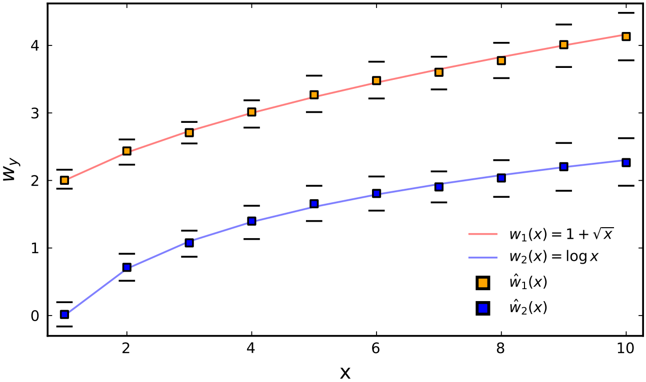
$$\begin{aligned} w_0(x) &= 0, \\ w_1(x) &= 1 + \sqrt{x}, \\ w_2(x) &= \log x. \end{aligned} \tag{27}$$

Population sample shares are calculated for $x = \{1, \ldots, 10\}$.

Compute $(\hat{w}_y^x)_{y=0}^2$ using discretization with $S = 1000$ *for each* x.

- ▷ indeterminacy problem irrelevant up to numerical error
- ▷ runtime about 0.22 seconds (on a 2016 laptop) per x

Figure 1: Simulation Results



Notes. Results based on 100 simulations with $S = 1,000$. Squares indicate mean estimates across simulations. Horizontal lines indicate corresponding 10% and 90% empirical percentiles. Coefficients are normalized s.t. $\hat{w}_0(x) = 0, \forall x$.